Development and validation of automatic ultrasound nerve block guidance for regional anesthesia using deep learning

Noam Suissa, B.Eng.

Department of Surgery

Division of Experimental Surgery

McGill University

Montreal, Quebec, Canada

June 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

MSc Experimental Surgery

Table of Contents

Abstract *5*

Abrégé *7*

Contribution of Authors and Acknowledgements *9*

Keywords *10*

Introduction *11*

Article 1 .... *13*

Abstract **14**

I. INTRODUCTION **15**

II. METHODS **18**

A. Data Collection 18

B. Labelling 18

C. Positive and Negative Images 18

D. Augmentations 19

E. Model architecture 22

F. Training 23

G. Testing 24

III. RESULTS **26**

IV. DISCUSSION **29**

REFERENCES **31**

Article 2 ... *33*

Abstract ... **35**

Introduction ... **37**

Methods ... **38**

Pixel-wise inter-labeler analysis ... 38

Survey-based inter- and intra-labeler analysis ... 42

Medical experience survey ... 44

Results ... **45**

Discussion ... **51**

Appendix I ... **55**

Supplemental Data ... **56**

Contributions ... **60**

References ... **61**

Literature Review ... *65*

AI in medicine ... **66**

AI in anesthesiology ... **70**

Education in anesthesia ... **74**

Medical Image Segmentation ... **77**

Overview of various blocks ... **90**

Risks and benefits of UGPNB ... **94**

Discussion & Conclusion ... *101*

References 104

Abstract

Background: Ultrasound-guided regional anesthesia (UGRA) has offered multiple advancements in the field of anesthesiology compared to traditional nerve stimulation techniques. However, nerve block and general sonoanatomy interpretation pose important challenges to novice anesthesiologists. This thesis will cover two articles which investigate deep learning (DL) techniques to highlight nerve regions in real time. An additional article will cover the inter- and intra-labeler variability among anesthesiologists' interpretations of target nerves in ultrasound (US) images.

Methods: A preliminary study was conducted on identifying the Transversus Abdominis Plane (TAP) using DL. The popular U-Net architecture was used to train a model on a total of 50,000 augmented positive images of the nerve and negative arbitrary US sonoanatomy. Ten anesthesiologists were recruited to label a test set (n=10) of TAP US images. Mean Dice scores were calculated for each of the ten anesthesiologists' labels of the test set. A global Dice score was calculated and represented the model's overall performance.

The second study on inter- and intra-labeler variability of nerve block target zones focused on establishing quantifiable links between subjective nerve target zone segmentation of still ultrasound images by anesthesiologists and objective similarity metrics. Ten anesthesiologists were recruited to label 70 US images: 7 nerve regions, 10 images for each region. Eight similarity coefficients, including the Dice score, were calculated for pairs of anesthesiologists' labels of US images. These scores were compared to the labeler's Yes/No blinded answers to whether they would insert a needle in their own or their colleague's labeled area. The same experiment was performed for two of the most experienced anesthesiologists in the cohort. Logistic regression coefficients and area under the curves (AUC) of receiver operating characteristic (ROC) curves were calculated for the following groups: All physicians, All nerves (APAN); All physicians, no Axillary (APnA); Experts, All nerves (EAN); Experts, no Axillary (EnA). Lastly, statistical data regarding the cohort's experience versus average agreement score was recorded.

Results: The overall Dice score for the preliminary study on automatic segmentation of the TAP region was 73.31%. In the expert comparison study, the Dice score ROC thresholds for APAN, APnA, EAN, and EnA were 34.44 (AUC: 0.72, p<<0.05), 81.35 (AUC: 0.61, p<<0.05), 64.67 (AUC: 0.65, p=0.0486), 59.63 (AUC: 0.90, p=0.1961), respectively.

Conclusion: The expert comparison study provided preliminary insight into establishing clinically relevant thresholds for important pixel-wise similarity metrics. The study investigating the TAP block resulted in a global Dice score of 73.31%, a score considered to be satisfactory, indicating that the model's segmentations of the region are in accordance with clinical standards.

Abrégé

Contexte : L'anesthésie régionale échoguidée (ARU) a permis de nombreuses avancées dans le domaine de l'anesthésiologie par rapport aux techniques traditionnelles de stimulation nerveuse. Cependant, l'interprétation du bloc nerveux et de la sonoanatomie générale pose d'importants défis aux anesthésistes novices. Cette thèse couvrira deux articles qui étudient les techniques d'apprentissage profond (DL) pour mettre en évidence les régions nerveuses en temps réel. Un article supplémentaire couvrira la variabilité inter- et intra-étiquette parmi les interprétations des anesthésistes des nerfs cibles dans les images échographiques (US).

Méthodes : Une étude préliminaire a été menée sur l'identification du plan transverse de l'abdomen (PTA) à l'aide de DL. L'architecture populaire U-Net a été utilisée pour former un modèle sur un total de 50,000 images positives augmentées du nerf et d'images échographiques arbitraires négatives. Dix anesthésistes ont été recrutés pour étiqueter un ensemble de dix images test US du PTA. Les scores de Dice moyens ont été calculés pour chacun des étiquetages de l'ensemble de test par les dix anesthésistes. Un score global de Dice a été calculé et représentait la performance globale du modèle. La deuxième étude sur la variabilité inter- et intra-étiquette des zones cibles du bloc nerveux visait à établir des liens quantifiables entre la segmentation subjective de la zone cible du bloc nerveux des images échographiques fixes par les anesthésistes et les mesures de similarité objectives. Dix anesthésistes ont été recrutés pour étiqueter 70 images US : 7 régions nerveuses, 10 images pour chaque région. Huit coefficients de similarité, dont le score de Dice, ont été calculés pour les paires d'étiquettes d'images US des anesthésistes. Ces scores ont été comparés aux réponses Oui/Non données en aveugle par les étiqueteurs pour savoir s'ils inséreraient une aiguille dans leur propre zone étiquetée ou dans celle de leur collègue. La même expérience a été réalisée pour deux des anesthésistes les plus expérimentés de la cohorte. Les coefficients de régression logistique et l'aire sous les courbes (AUC) des courbes ROC (receiver operating characteristic - fonction d'efficacité du récepteur) ont été calculés pour les groupes suivants : Tous les médecins, tous les nerfs (APAN) ; Tous les médecins, pas d'Axillaire (APnA) ; Experts, tous les nerfs (EAN) ; Experts, pas d'Axillaire (EnA). Enfin, des données statistiques concernant l'expérience de la cohorte par rapport au score d'accord moyen ont été enregistrées.

Résultats : Le score global de Dice pour l'étude préliminaire sur la segmentation automatique de la région TAP était de 73,31 %. Dans l'étude de comparaison d'experts, les seuils ROC du score de Dice pour APAN, APnA, EAN et EnA étaient respectivement de 34,44 (AUC : 0,72, $p<0,05$), 81,35 (AUC : 0,61, $p<0,05$), 64,67 (AUC : 0,65, $p=0,0486$), 59,63 (AUC : 0,90, $p=0,1961$).

Conclusion : L'étude de comparaison d'experts a fourni un aperçu préliminaire de l'établissement de seuils cliniquement pertinents pour d'importantes mesures de similarité au niveau du pixel. L'étude du bloc TAP a abouti à un score global de Dice de 73,31 %, un score considéré comme satisfaisant, indiquant que les segmentations de la région par le modèle sont conformes aux normes cliniques.

Contribution of Authors and Acknowledgements

Noam Suissa is the sole author of this entire thesis. Furthermore, Noam Suissa is first author in Articles 1 and 2; all other co-authors are stated at the beginning of each article.

A special thank you to all anesthesiologists who participated in the studies. Your efforts have shed light on important findings in both fields of anesthesiology and machine learning.

Keywords

Abbreviations	**Definition**
AI	Artificial Intelligence
CNN	Convolutional neural network
DL	Deep Learning
FN	Femoral nerve
ML	Machine learning
MN_UN	Median nerve and Ulnar nerve
RN	Radial nerve
TAP	Transversus abdominis plane
UGRA	Ultrasound-guided regional anesthesia
UGPNB	Ultrasound-guided peripheral nerve blocks
US	Ultrasound
SN	Sciatic nerve

Introduction

Peripheral regional anesthesia, or more commonly known as nerve blocks, have been a common procedure to induce analgesia in patients perioperatively (Henderson et al., 2016). In the last several decades, ultrasound has gained traction in the field of regional anesthesia, effectively switching the procedural standard from nerve stimulation and landmark-based nerve blocks to ones performed using ultrasound guidance (Marhofer et al., 2005). However, novice anesthesiologists face many challenges with this new technique such as sonoanatomy interpretation, needle guidance; all of which, if not careful, can lead to inadvertent organ damage or anesthetic toxicity (Scherrer et al., 2013). In order to address these issues, multiple machine and deep learning solutions have been developed in recent years with the goal of automatically highlighting or surrounding important anatomical landmarks such as nerves or vessels in ultrasound images in real time. With these technologies, researchers hope that they can be used not only as a real-time decision support tool but can also be used towards training novice anesthesiologists in better recognizing complex sonoanatomy.

In this manuscript-based thesis, two submitted articles to high impact-factor journals are presented. The first article, submitted and accepted to the IEEE journal in May 2023, explored the efficacy of highlighting the transversus abdominis plane (TAP) nerve region in US images using deep convolutional neural networks (CNN) (Suissa et al., 2023, In Press). The network was trained on 50,000 positive and negative (i.e. containing/missing the TAP nerve) US images and evaluated against 10 images that have been labeled by 10 anesthesiologists. The study is presented in full in the chapter titled Article 1.

The second article, submitted to the British Journal of Anesthesia (BJA) in May 2023 and is pending approval, highlights the inter- and intra-labeler variability amongst anesthesiologists' labels of nerve regions in US images (Suissa et al., 2023, Pending Approval). Furthermore, the study establishes an optimal threshold that links an anesthesiologist's subjective label to an objective pixel-wise comparison metric such as the Dice coefficient. The objective of this study was to help machine learning researchers who are performing US nerve segmentation relate their model's scores to a clinically relevant and statistically-backed threshold. The entire study can be found in the chapter titled Article 2.

Lastly, an in-depth literature review covering over 70 academic articles of the relevant literature in artificial intelligence and anesthesia is presented in the Literature Review section.

The thesis ends in the Discussion and Conclusion chapter, which outlines the potential impact that these studies can bring to the fields of machine learning and anesthesia. Moreover, the limitations and future directions of this project are also discussed.

Article 1

This article has been submitted and accepted to the IEEE EMBC 2023 conference as of April 2023.

Original format of the submission has been preserved.

Utilizing Deep Learning to Identify an Ultrasound-guided Nerve Block Target Zone

Noam Suissa, Sean D Jeffries, Kevin Song, Samer Botros, Thomas M Hemmerling

Research supported by funds from ITAG laboratory, McGill.

N.S and S.D.J. authors are with the Department of Experimental Surgery, McGill University, Montreal, Quebec, Canada.

T.M.H is with the Dept. of Anesthesiology and Division of Experimental Surgery, McGill University, Montreal, Quebec, Canada. (corresponding author phone: 5149341934 ext 43677; e-mail: thomas.hemmerling@ mcgill.ca).

Abstract – Ultrasound guided nerve blocks are increasingly being used in perioperative care as a means of safely delivering analgesia. Unfortunately, identifying nerves in ultrasound images presents a challenging task for novice anesthesiologists. Drawing from online resources, here we attempted to address this issue by developing a deep learning algorithm capable of automatically identifying the transversus abdominis plane region in ultrasound images. Training of our dataset was done using the U-Net architecture and artificial augmentation was done to optimize our training dataset. The Dice score coefficient was used to evaluate our model, with further evaluation against a test set composed of manually drawn labels from a pool of (n=10) expert anesthesiologists.

Across all labelers the model achieved a global Dice score of 73.31% over the entire test set. These preliminary results highlight the potential effectiveness of this model as a future ultrasound decision support system in the field of anesthesia.

I. INTRODUCTION

Peripheral nerve blocks are important facets of any procedure involving regional and local anesthesia. The nerve block is a technique that involves the desensitization of a general area of the body, driven through a targeted injection of anesthetic into the surrounding tissue of a nerve that directly controls the area of interest (1). These blocks allow for the management of chronic pain following surgery, therapeutic control of acute pain, origin diagnoses of pain, and even have prognostic implications for potential surgical interventions (2). Effects can be brief or extended over a long period of time depending on procedural needs (1). As such, peripheral nerve blocks have quickly become a popular course of action in any anesthetist's treatment plan (2).

The rational for nerve block analgesia's rising popularity is primarily twofold – the avoidance of complications related to general anesthesia and the minimization of opioid use, all while still delivering extended pain relief (2). Moreover, with the recent emergence of widely available ultrasound imaging techniques, the efficacy and safety profiles related to nerve block performance have risen drastically (3). Unfortunately, there still currently exists a lack of knowledge depth in sonoanatomy for new graduates in anesthesiology without fellowship specialty training (4). One of the most critical components of ultrasound (US) -guided peripheral nerve blocks is the accurate detection of a nerve region within an ultrasound image (3). Taking into consideration anatomical variability between patients, classifying various types of tissue in a noisy, greyscale ultrasound feed can be a daunting task to the uninitiated (4).

The use of artificial intelligence-based image recognition software could prove useful in solving this problem as a tool in both teaching and procedural guidance. Similar solutions have proven useful in other medical specialties that rely on imaging as a crucial part of standard practice; for example, deep learning and quantitative modelling in the fields of radiology and oncology are now seen as potentially effective tools for use in early detection, diagnosis, and prognosis of disease (5). In this paper we investigate whether a software solution could also possibly aid physicians in performing ultrasound-guided peripheral nerve block procedures in a more precise and effective manner. The transversus abdominis plane (TAP) block targets anterior rami of spinal nerves, typically from the seventh thoracic to first lumbar spinal nerves (1). Although the actual rami cannot be visualized using US-guidance, the plane between the respective abdominal wall muscles in which these nerve endings are located can be visualized (1). The plane block target area is located between the internal oblique muscle and the transversus abdominis muscle (Figure 1).

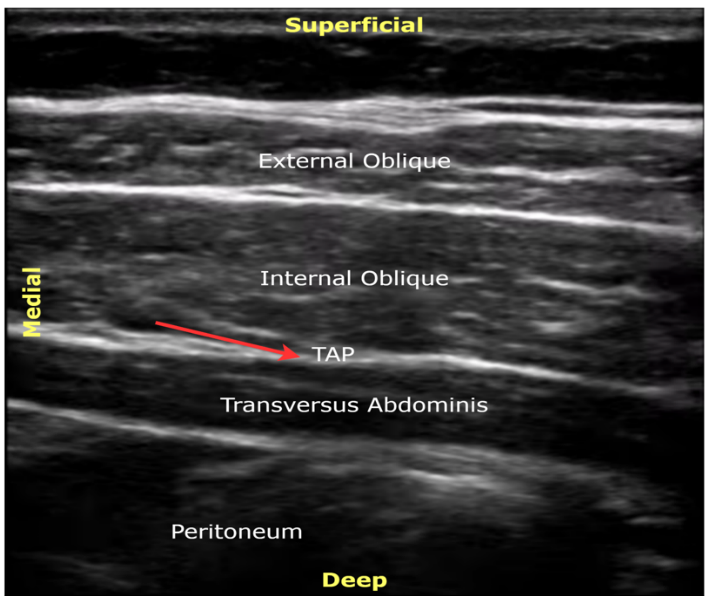

Figure 1. Sonoanatomy of the TAP Block Target Region

The goal of this project was the construction of a deep learning convolutional neural network (CNN) capable of assisting physicians in the detection of a specific nerve block region with a high degree of accuracy. Specifically, this tool was trained to guide anesthesiologists throughout the duration of a TAP block procedure, serving as a novel description of a future decision support system. The model described in this study could further be expanded upon to focus on other nerve regions and stands as a proof of concept to increase the future efficacy and safety profiles of procedures within the field of anesthesiology.

II. METHODS

A. Data Collection

Still ultrasound images were sourced from various TAP block YouTube videos (Alphabet Inc.®, Mountain View, CA). The primary objective when taking still frames from each video was to ensure that all frames had subtle to large differences in appearance, allowing for the final dataset to be composed of unique images. A total of 554 images were sourced from 44 YouTube videos, with 544 of those comprising the training set and the remainder forming the test set. Furthermore, the images were resized to 128x128 pixels for faster training and more efficient computing.

B. Labelling

The labeling of the training and test sets was performed by anesthesiologist T.H. An additional nine anesthesiologists who were familiar with the TAP block and performed several monthly, ranging from 2-80 blocks/month, labeled the same images within the test set. The web-based labeling software LabelBox (Labelbox Inc.®, San Francisco, CA) was used to collect the manually drawn labels for the training and test sets.

C. Positive and Negative Images

The 554 ultrasound images collected containing the TAP plane were used to comprise the positive dataset. Naturally, for a model to be successful in predicting a specific region in an image, it must also avoid making a prediction when the region is not present in the image. To achieve this, a training dataset composed of positive and negative cases is required, with the latter significantly reducing false positive predictions (6). Using the same methodologies for collecting the positive cases, we searched through numerous YouTube videos containing any human ultrasound images, excluding that of fetal ultrasound. The reasoning behind taking any ultrasound images other than the TAP region as a negative is to train the model to only make predictions on the region of interest and nowhere else, effectively minimizing chances of a false positive (6). A total of 908 negative images were collected, consisting of benign, malignant, and tumor free breast ultrasounds, thyroid ultrasounds, as well as femoral and popliteal ultrasounds.

D. Augmentations

Data augmentation is a technique used in machine and deep learning to artificially increase the size of the training dataset. Instead of having to manually search for new unique images, clever techniques are used to introduce subtle to large differences in the source images, producing visually distinct ones altogether. The image augmentations used in this study follow those presented in Smistad et al., namely: flipping, rotation, gamma intensification, elastic deformation, and gaussian shadow (7). Figure 1 demonstrates the augmentations applied to the original images and their masks.

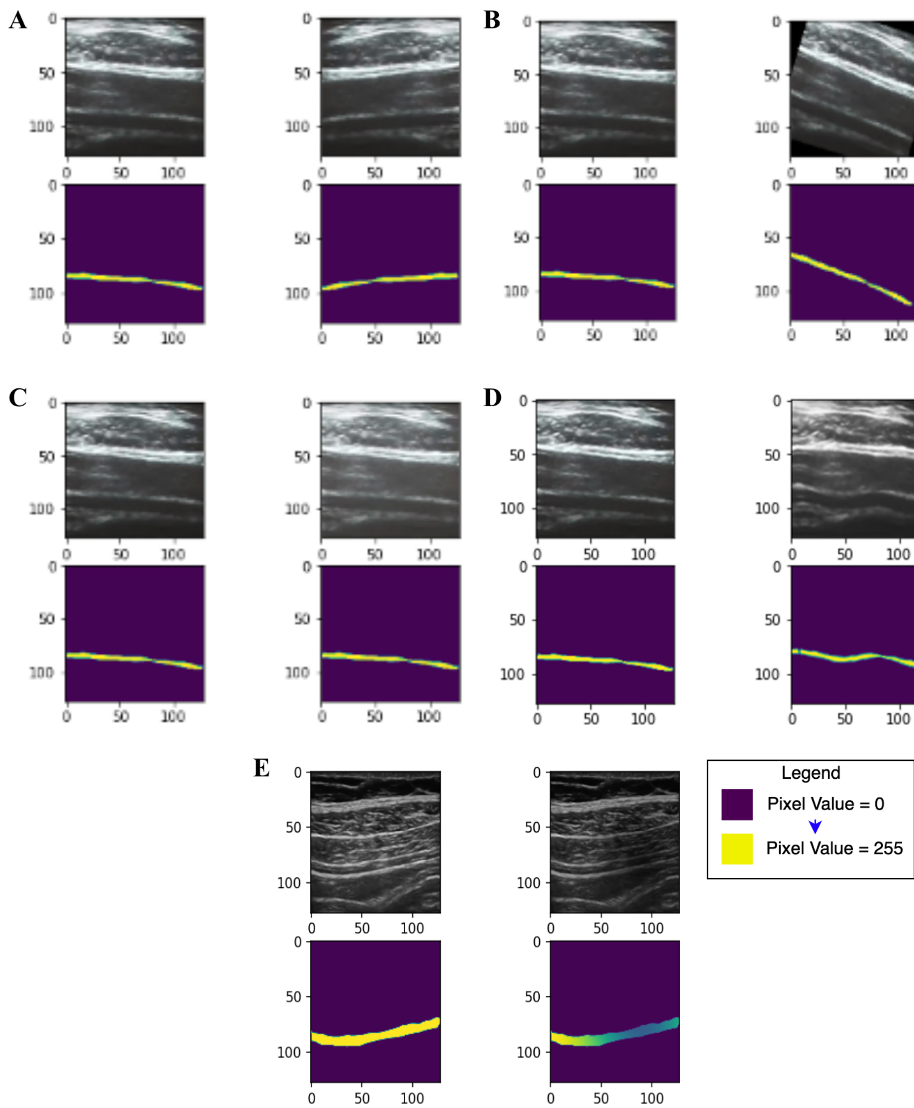

Figure 2. Examples of Image and Subsequent Mask Prior to Specific Augmentation and Following Augmentation.

(A) Flipping of Image. (B) Rotation of Image. (C) Image Gamma Intensification. (D) Image Elastic Deformation (E) Application of Gaussian Shadow to Image.

Flipping consisted of flipping all the image-mask pairs horizontally (Figure 1A). Image rotation was achieved by rotating the image-mask pairs by a random angle, in degrees, within the range [-10, 10] (Figure 1B). By adjusting the gamma intensity on our ultrasound images, it is possible to mimic the gain on real ultrasound machines (7). We achieved this by raising each pixel value by a randomly chosen gamma coefficient from a predefined range (Figure 1C). Elastic deformations are another great way to augment image data when dealing with patient populations, as the resulting images can help account for variations in sonoanatomy across patients (8). We elastically deformed our images here by displacing pixels in the original image according to a randomly generated displacement field (Figure 1D). Finally, it was important to recognize that pockets of air in surface tissue often block ultrasonic waves from reaching the underlying tissue of interest, creating acoustic shadows (7). To simulate a shadow programmatically, we generated a 2-dimensional gaussian shadow overlay by convolving it over the original image (Figure 1E). It is worth noting that because the shadows are never truly opaque, the shadows never completely hide portions of the label. Moreover, with the labels consisting of Boolean matrices, the resulting label would remain unchanged; any positive pixel value becomes "True". This was deemed to be acceptable for two reasons: first, as mentioned above, the nerve is never completely hidden, and second, we still achieve the goal of augmenting the data by introducing slight changes in the original images.

Our goal was to train the model on approximately 25,000 images from the base dataset of 544 images. In order to achieve this, we first begin by passing each image to the augmentation script where they would have a 50% chance of undergoing an augmentation in the same order of augmentations presented above. To this end, a particular image had a chance of being transformed five times. Over the entire dataset, we observe that on average between two and three augmentations were applied to the images. In the case where an image was not chosen for the application of an augmentation, a default and random gaussian shadow was applied. Finally, to reach our desired dataset size, this process would be repeated on the original dataset's images until we had 25,000 images. This same process was applied to the negative image dataset for a total of 50,000 images.

E. Model architecture

The U-Net is a precise image segmentation CNN architecture developed by Olaf Ronneberger et al. at the University of Freiburg in Germany (9). It was selected as an ideal candidate to train our model on, as the architecture is designed to tackle image segmentation tasks and performs well with limited datasets (9). The U-Net used in this project varies slightly from the original one proposed by researchers at the university of Freiburg (Figure 2). Excluding the difference in the input size, their model crops every contracting step's output before concatenation with the expanding layers. Ultimately, this yields a smaller image size than the original input. To avoid this, we only concatenate and skip the cropping step. Furthermore, the model performs successive max pooling operations until we extracted 256 8x8 pixel feature sets. Overall, our model contained a total of 1,941,105 trainable parameters.

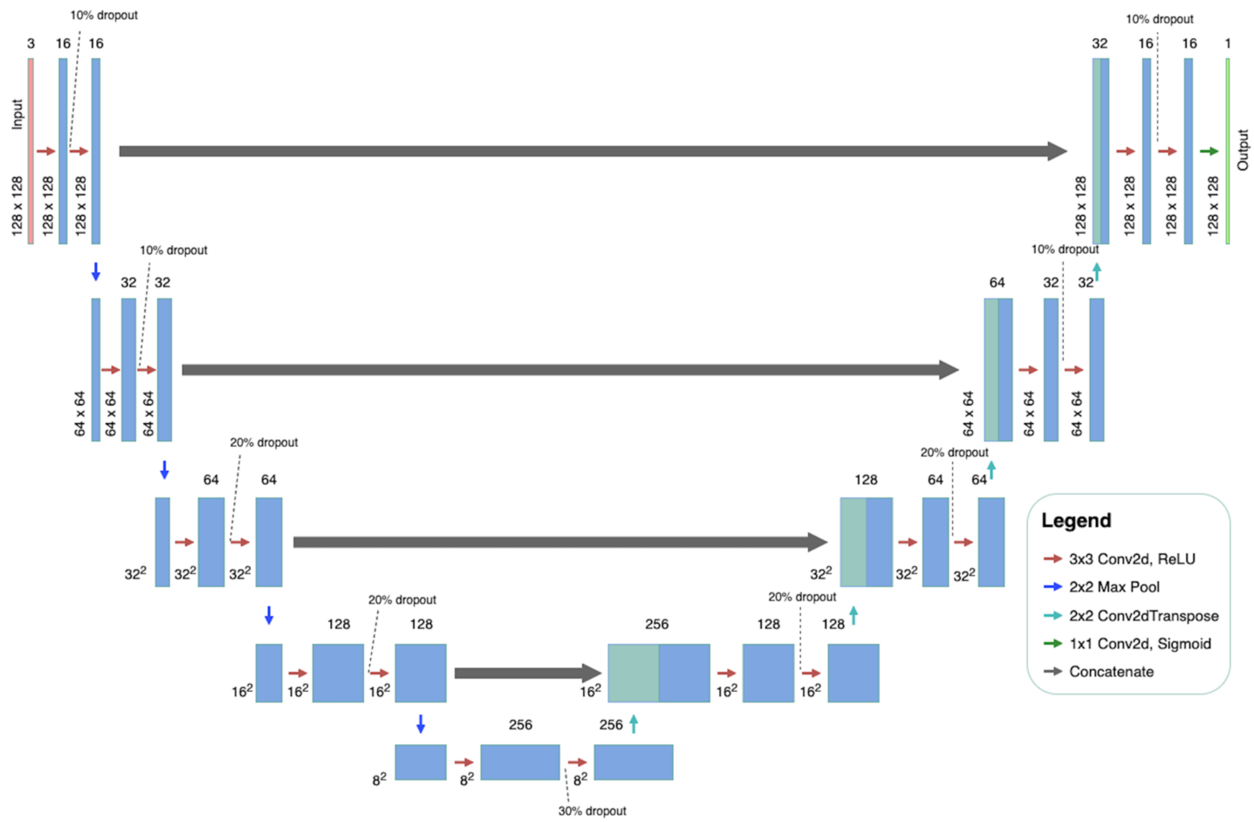

Figure 3. The U-Net Architecture Utilized in this Model (9)

F. Training

The model was trained for 10 epochs with a batch size 16 on an EVGA Geforce Nvidia RTX 3090. Due to memory limitation in regard to our dataset size of ~50,000 images (positive and negative cases), the datasets were saved to separate HDF5 files. Upon training a generator would progressively retrieve eight images from each set in a shuffled order for a total batch size of 8 images. Furthermore, 10-fold cross validation is used to minimise bias and loss even further. Lastly, training and the implementation of our network was done using the Keras framework packaged with Tensorflow 2.5.0 (Alphabet Inc.®, Mountain View, CA).

G. Testing

Ten anesthesiologists from four different centers labeled ten test images of TAP nerves to compare with our model. The trained model would then produce predictions of the TAP nerve for each image. The Dice score (Equation 1), a metric commonly used in deep learning segmentation tasks, compares two labels' pixel-wise overlap accuracy and produces values ranging from 0 to 1. Zero indicates that the two regions do not overlap at all while a score of one represents a perfect match. The calculation requires four distinct values: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). With that said, the Dice score was recorded for each prediction-physician label pair. The average Dice score across all ten images (N) for that specific labeler (L) was then recorded, and the final metric representing the model's overall performance was the global average of each physician's mean scores (Equation 2).

$$Dice(x, y) = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{1}$$

$$Global\ Dice = \frac{1}{L} \sum_{y=1}^{L} \frac{1}{N} \sum_{x=1}^{N} Dice(x, y) \times 100 \tag{2}$$

Whereas each prediction produces a heat map with pixel values ranging from 0 to 1, the Dice score only works when comparing boolean values. In order to achieve this, a cutoff threshold for the prediction's pixel values must be obtained. This will have set the pixel values that exceed the threshold to True and the remaining to False. This process provided counts for the four required values. In all, we face an optimization problem; finding the cutoff threshold which maximizes the overall Dice score. This was achieved by plotting the global Dice score for each 0.01 increment of pixel value cutoff. The cutoff which then produced the highest global dice score was selected for future inferencing of the TAP region.

III. RESULTS

Training times were relatively long to complete with each epoch taking an average of 87 seconds, for a total training time of four hours and fifty minutes. The model achieved a minimal validation loss of 0.018. The evaluated cutoff using our methods was calculated to be 0.51. Using this threshold, we obtained a maximum global dice score of 73.31% (Figure 3).

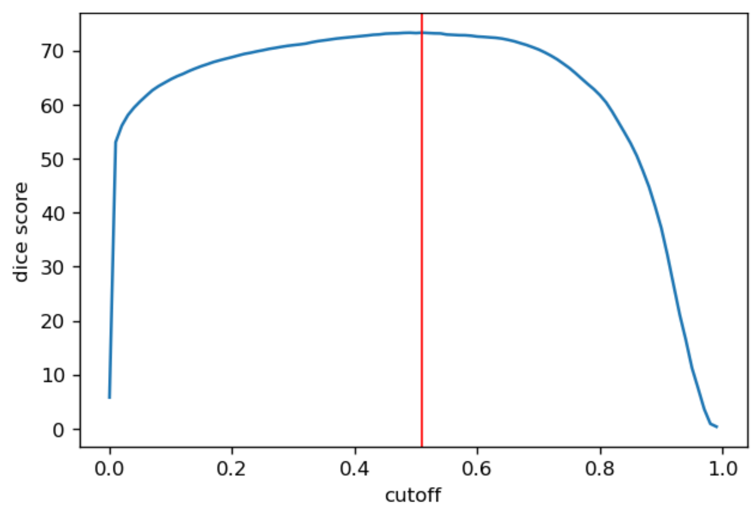

Figure 4. Global Dice Score Rate of Change Curve

Labeler	**Mean Dice Score (SD)**
Labeler 1	69.11 (9.43)
Labeler 2	81.76 (3.55)
Labeler 3	81.11 (5.64)
Labeler 4	78.82 (6.54)
Labeler 5	62.23 (10.69)
Labeler 6	78.85 (4.52)
Labeler 7	72.27 (7.01)
Labeler 8	67.29 (12.47)
Labeler 9	83.92 (3.22)
Labeler 10	56.71 (10.69)
Global Dice	**73.31**

Table 1 – Mean Dice Scores Across All Labelers

Table 1 includes all the individual mean scores of each physician across the ten test images from which the overall average was calculated. After having evaluated the cutoff, the predictions could then be applied over the test images. An example of this final prediction is demonstrated in Figure 4.

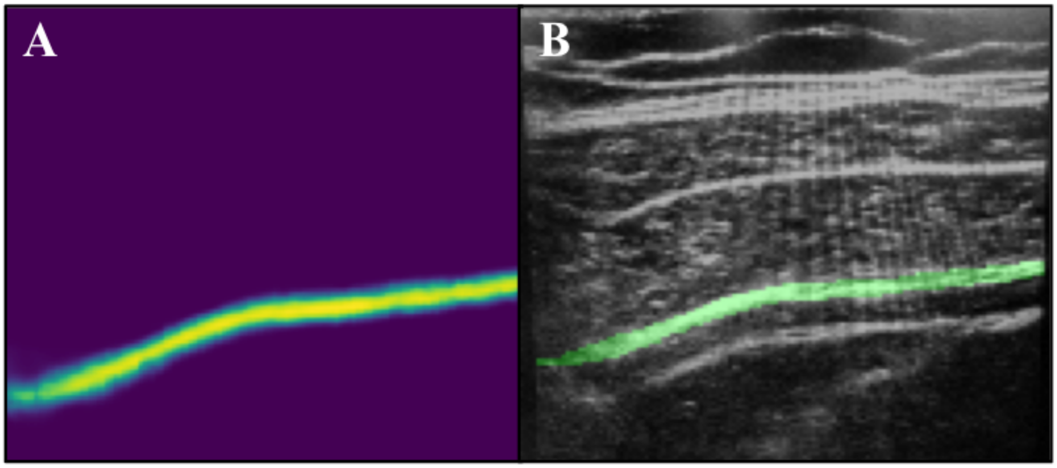

Figure 5. Transverse Abdominis Model Prediction on a Randomly Selected Image

IV. DISCUSSION

Recent advances in artificial intelligence have the potential to impact several fields of medicine by improving the accuracy and efficacy of various esoteric tasks (5). While these programs have proved useful in many imaging heavy disciplines such as radiology, more nuanced approaches in other departments warrant further study (7). Here, we attempted to combine previous neural network imaging research to the department of anesthesiology, particularly that of ultrasound guided nerve blocks. To prove the efficacy of this work, our laboratory chose to investigate whether an artificial intelligence-based segmentation model could effectively predict the transversus abdominis plane region in a variety of ultrasound images.

The trained CNN presented in this work was successful at making predictions on a set of 10 still ultrasound images, with a maximum global Dice score of 73.31%. This represents a satisfactory score when compared to other machine learning problems in the field, however, no absolute threshold value pertaining to clinical relevancy has yet been established.

The main limitation of this study was the variations between labelers' manual segmentations within the test set, known to affect the overall Dice score of models, however, the relationship and magnitude of this effect has yet to be thoroughly studied (5). Additional limitations include the size of the test set (n = 10) when compared to that of the training set (n = 50,000), restricted by expert availability and time required to complete the labelling tasks. Finally, our model was trained on the first iteration of the U-Net architecture which has previously proven useful in the detection of nerves, nevertheless, newer versions are now available and are known to perform better (9).

Our program described here serves as a proof of concept for prospective clinical decision support systems. This novel program appears to function with a satisfactory degree of accuracy in detection of the transversus abdominis plane. In the future, we plan on expanding our algorithm to detect a wide range of nerves essential in the most utilized nerve block procedures. We also plan on exploring the role of hyperparameter optimization and model architecture selection in the optimization of our models segmentation accuracy. Finally, once enough data has been accumulated and labelled, we intend to shift our program from a still-image-based detection software to a real-time, live, nerve detection program. If successful, this will serve as a first step in the creation of a decision support system for clinicians to use in a wide variety of ultrasound imaging-based nerve block procedures. Such a tool could greatly improve the accuracy, effectiveness, and efficiency of medical care from an anesthetic perspective.

REFERENCES

[1] Chang, A., Dua, A., Singh, K., & White, B. A. (2022). Peripheral Nerve Blocks. In StatPearls. StatPearls Publishing.

[2] Wiederhold, B. D., Garmon, E. H., Peterson, E., Stevens, J. B., & O'Rourke , M. C. (2022). Nerve Block Anesthesia. In StatPearls. StatPearls Publishing.

[3] Warman, P., & Nicholls, B. (2009). Ultrasound-guided nerve blocks: efficacy and safety. Best practice & research. Clinical anaesthesiology, 23(3), 313–326. https://doi.org/10.1016/j.bpa.2009.02.004

[4] Gharapetian, A., Chung, F., Wong, D., & Wong, J. (2015). Perioperative fellowship curricula in anesthesiology: a systematic review. Canadian journal of anaesthesia = Journal canadien d'anesthesie, 62(4), 403–412. https://doi.org/10.1007/s12630-014- 0299-2

[5] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future healthcare journal, 6(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

[6] Gao, L., He, Y., Sun, X., Jia, X., & Zhang, B. (2019). Incorporating Negative Sample Training for Ship Detection Based on Deep Learning. Sensors (Basel, Switzerland), 19(3), 684. https://doi.org/10.3390/s19030684

[7] Smistad, E., Johansen, K. F., Iversen, D. H., & Reinertsen, I. (2018). Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks. Journal of medical imaging (Bellingham, Wash.), 5(4), 044004. https://doi.org/10.1117/1.JMI.5.4.044004

[8] Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: fast and flexible image augmentations. Information, 11(2), 125.

[9] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation.
In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Article 2

This article was submitted to the British Journal of Anesthesia (BJA) on May 30th, 2023. Original format has been preserved, however, figures and tables have been added in line with the text.

Quantifying ultrasound evaluation metric significance in medical image segmentation: comparison of expert evaluation

Noam Suissa[1], Sean D. Jeffries[1], Jose Luis Ramirez-GarciaLuna[1], Joshua Morse[1], *Thomas M. Hemmerling[1,2]

1. Department of Experimental Surgery, McGill University Health Center, Montréal, Canada

2. Department of Anesthesia, McGill University, Montréal, Canada

*** Corresponding Author**

Thomas M. Hemmerling, MD, Department of Anesthesiology and Experimental Surgery

McGill University and Montreal General Hospital 1650 Cedar Ave, T8-108

Montreal, Quebec H3G 1A4, Canada (e-mail: Thomas.Hemmerling@mcgill.ca).

Conflicts of interest: The authors declare no conflicts of interest

Funding: The authors have no sources of funding to declare for this manuscript

Keywords – Ultrasound, Nerve Block, Regional Anesthesia, Sonoanatomy, Dice Score, Machine Learning, Segmentation

Abstract

Background: Ultrasound-guided regional anesthesia (UGRA) offers unprecedented guidance for anesthesiologists performing nerve blocks. With continuous visualization of target sonoanatomy, anesthesiologists must skillfully interpret live feeds to avoid trauma from needle placement. This study attempts to quantify agreement between anesthesiologists' interpretation of target nerves in common UGRA blocks.

Methods: Ten anesthesiologists of various ranks and expertise labelled a total of 70 ultrasound images coming from seven different nerve regions. One randomly selected image per region was used to produce a set of 70 labelled images. A set of eight comparison metrics were computed for every permutation of pairs of anesthesiologists' labels. Every physician answered yes/no questions as to whether they would guide the needle to colleagues labelled areas while being blinded. This experiment was then repeated for two experts in the cohort. The data gathered was split into four groups: All Physicians All Nerves (APAN), All Physicians no Axillary (APnA), Experts All Nerves (EAN), and Experts no Axillary (EnA). Binary logistic regressions, ROC curves, ICC, and Cohen's Kappa were performed on the date set. Lastly, a logistic regression was performed on the cohort's experience data.

Results: Dice score ROC thresholds for APAN, APnA, EAN, and EnA were 34.44 (AUC: 0.72, $p \ll 0.05$), 81.35 (AUC: 0.61, $p \ll 0.05$), 64.67 (AUC: 0.65, $p=0.0486$), 59.63 (AUC: 0.90, $p=0.1961$), respectively. Years of professional experience post-residency had an inversely proportional relationship to the average labeling score with a factor of -0.571 ($p=0.045$).

Conclusion: Relatively low pixel-wise comparison metric scores correlate to a substantial agreement on nerve location between raters. This study hopes to establish a benchmark for future

machine and deep learning studies working on medical segmentation tasks and serve as a guide for evaluating future model performance and clinical relevance.

Introduction

The advent of ultrasound-guided peripheral nerve blocks (UGPNB) has revolutionized the practice of regional anesthesia.[1] Compared to traditional surface landmark techniques that anesthesiologists use to gauge the location of a nerve, UGPNB offers continuous visualization of the nerve and surrounding sonoanatomy.[2] This live visualization allows for a more precise spread of the injectate, and as a consequence, faster sensory onset and an overall increase in blocking success rates.[3] In a 2016 study conducted by Henderson and colleagues, the authors point out that despite the many advantages of UGPNB, certain procedures can still pose unique challenges.[4] According to the authors, acoustic artifacts that are produced by the ultrasound (US) probe's emitted sound waves can produce brighter or darker regions in the US image.[4] This in turn may confuse anesthesiologists in their interpretations and later cause complications.[4] Machine learning (ML) and deep learning (DL) articles in 2017 have collectively accounted for over 50% of publications in the fields of computed tomography, magnetic resonance imaging, and US imaging.[5] At present, US imaging constitutes 5% of published articles.[5] The segmentation of anatomical regions in medical US images has consistently been a vital and significant process in the development of machine learning models, particularly for tasks related to diagnostics or procedures.[5] Multiple studies and experiments to automatically highlight nerve regions using ML and DL techniques have been published in recent years to address these issues.[6-10] However, these automated UGPNB solutions fail to provide insight into the clinical relevance of their chosen accuracy metric. In this study, we aimed to provide a statistically backed benchmark for UGPNB segmentation predictions by looking at the inter- and intra-labeler variability of multiple anesthetists' manual segmentation of several ultrasound images. The statistics gained from this study will allow future predictive algorithms in UGPNB to better interpret their results in the frame of clinical relevance.

Methods

Seven nerve regions among six widely performed nerve blocks were selected for this experiment. Using the popular labeling software tool Labelbox (Labelbox Inc.®, San Francisco, CA), ten volunteer anesthesiologists of various expertise from four hospitals were asked to segment the same ten images for each selected nerve block. Images were sourced from 246 YouTube videos, with authors only selecting frames for inclusion that did not contain US software-related artifacts such as depth of scan and virtual measurements. Each image in a set was chosen to feature differences in anatomical dispositions so as not to include similar-looking images. The regions in question were as follows: the transversus abdominis plane (TAP block); Pecs I (PECS I block); Posterior Rectus Sheath (Rectus Sheath block); Sciatic Nerve (SN, Popliteal Fossa block); Femoral Nerve (FN, Femoral nerve block); Radial Nerve (RN) as well as the Median and Ulnar nerves (MN_UN). The RN, MN, and UN were all contained in the Axillary Brachial Plexus (BP) block. The Femoral Nerve (FN) dataset was uniquely sourced from the femoral_nerve_block_computer_vision GitHub repository.[11]

Pixel-wise inter-labeler analysis

Once all images were labeled, they were then exported from Labelbox for statistical analysis. When it comes to image segmentation, particularly for machine and deep learning applications, several metrics exist that can measure the pixel-wise overlaps of two areas in an image.[12,13] In an article published by Hicks and colleagues, the importance of including evaluation metrics for medical artificial intelligence applications is discussed.[14] The study presents 8 metrics, stemming from four distinct diagnostic accuracy parameters: true positive rate (TP), true negative rate (TN), false positive rate (FP), and false negative rate (FN).[14] However, while the paper presents ways to calculate agreement between an AI's prediction and a manual label, they are equally capable of being used to determine the similarities between two manually drawn segmentations of a region in an image. With that said, the order of the pair being compared is important as the definition of FP and FN can change. Therefore, when comparing two experts' labels, we defined which among the two was considered the "prediction" and "the ground truth". In this regard, pairs of experts were labeled *expert1-expert2* throughout our experiment to represent the ground truth and prediction, respectively.

Eight metrics were used to compare the pair of experts' labels' similarities, and each provided unique insights into the inter-labeler agreement.

Accuracy (eq. 1) refers to the ratio between correctly classified samples (pixels) and the total number of pixels in the images. This metric ranges from 0 to 1, implying no or complete overlap between the two areas.

$$\frac{TP+TN}{TP+FP+TN+FN} \quad \text{Eq. 1}$$

Sensitivity, Recall, or True Positive Rate (eq. 2) represents the rate of correctly classified positive pixels and takes the ratio between correctly predicted positive pixels and all those belonging to the ground truth. A sensitivity value of 0 means that the prediction completely missed the ground truth whereas a value of 1 indicates a perfect overlap.

$$\frac{TP}{TP+FN} \quad \text{Eq. 2}$$

Specificity (eq. 3) represents the rate of negative pixels correctly classified and comprises the ratio between correctly predicted negative pixels and all pixels not belonging to the ground truth. A value of 1 indicates a perfect prediction of the negative class while a value of 0 represents a complete miss.

$$\frac{TN}{TN + FP} \qquad \text{Eq. 3}$$

The Positive Predictive Value (PPV) or precision (eq. 4) represents the proportion of retrieved pixels that are relevant. It is calculated by taking the ratio between the correctly predicted positive pixels and all positively predicted pixels. A score of 1 represents a perfect prediction of the positive class.

$$\frac{TP}{TP + FP} \qquad \text{Eq. 4}$$

Similarly, the Negative Predictive Value (NPV) (eq. 5) represents the proportion of retrieved pixels that belong to the negative class. A score of 1 indicates a perfect prediction of the negative class.

$$\frac{TN}{TN + FN} \qquad \text{Eq. 5}$$

The F1 Score or Dice Coefficient (eq. 6) is the harmonic mean of precision and recall and is not class symmetric meaning that the definition of which class of pixels are positive and negative yields two different scores. A Dice score of 1 represents a perfect overlap between the ground truth and the prediction while a score of 0 indicates a complete miss between the two.

$$\frac{2 \times TP}{2 \times TP + FP + FN} \qquad \text{Eq. 6}$$

Matthews Correlation Coefficient (MCC) (eq. 7) is an excellent metric for datasets with imbalanced classes. The score ranges from -1 to 1 with 1 representing a perfect prediction, 0 indicating a prediction no better than a random one, and -1 being a total miss of the ground truth.

$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad \text{Eq. 7}$$

Finally, the Threat Score (TS) (eq.8) is a metric that is particularly sensitive to correct predictions of rare positive events and is calculated by taking the ratio between correct positive pixels and all incorrect predictions as well as the correct positive ones. The score ranges from 0 to 1; 1 indicating a perfect match.

$$\frac{TP}{TP + FN + FP} \quad \text{Eq. 8}$$

Once these metrics were established, the scores for each pair of labelers' segmentations over the entire set of images was calculated. Moreover, since certain scores are not labeler-symmetric and the definition of FP and FN could change based on the comparison order of labelers, the scores for every permutation of labeler pairs were also recorded. With ten labelers, each image produced a score table consisting of 90 rows (n=10P2) and eight columns per row for a total of 720 data points. In addition, 10 images per region and seven regions produce 50,400 data points. To condense this data, one large table containing the metric scores of a randomly selected image from each block was formed, which represented the values used in the following experiments. Receiver operating characteristic (ROC) curves were then produced from the binary logistic regressions between the dice scores and the yes/no results from the following experiment to obtain the optimal dice score thresholds.

Survey-based inter- and intra-labeler analysis

For this experiment, all manually drawn labels were overlayed onto their respective images producing a green translucent area over the block region (Figure 1, Supplemental Data) and passed to a modified version of an open-source labeling software called Swipe-Labeler published by Jenessa and colleagues[15]. This software is a simple web app that allows users to categorize images one by one quickly and efficiently. Before the start of the experiment, the rater was told that the app will present a series of labels from other experts and was not explicitly told that each region contains their own labels as well.

First, the user was asked to input their name to allow us to group their answers into their own folder. Next, the experiment begins by displaying a single image at a time of an expert's label with the question above reading: *Regarding the <nerve region>, would you guide the needle to the highlighted area to block this region?* where *<nerve region>* was replaced by the block type the displayed ultrasound image depicts. The user was then required to click on a yes or no button below the image which then sends the image file into its respective yes or no folder within that user's custom folder. Each image file was named in the following way: *<image number>* *<image_unique_ id>_<region>_<labeler>.png* for easy tracking later. Once all 70 images had been categorized by a rater, the results were transcribed into discernable data (Supplemental Data - Table 1). Furthermore, as seen in Supplemental Data Table 1, the highlighted cell in the Labelers column indicates that the scores in the table were retrieved from that rater and, to that end, the respective row represents the intra-labeler scores. Once every rater's scores were recorded, a summary table containing the sum of each respective cell and the average score per labeler was produced (Supplemental Data - Table 2). This effectively transformed the data from nominal to continuous, allowing for a more robust analysis and implementation of descriptive statistics. Moreover, an Intraclass Correlation Coefficient (ICC) test was performed to quantify the intra- and inter-labeler variability using this table. Additionally, these scores were then matched to their respective *expert1-expert2* row in the pairwise metric table, enabling the calculation of binary logistic regressions. The yes/no data were added as follows: expert1-expert2's row received expert2's yes/no answer from expert1's label. Given that this experiment only displayed a randomly chosen image of ten for each block, these scores were only be appended to one pairwise metric table corresponding to the scores for that image of that specific block.

Another experiment consisted of having the two most experienced physicians in the group, defined by years of experience, perform the survey together in a collaborative fashion. This process was conducted virtually by two moderators, [N.S] and [S.J]. Once an image was presented, the physicians sent a private message to one of the moderators with the answer to the displayed question (i.e., Yes or No). A follow-up question was then asked by one of the moderators to rate how much the anesthesiologists agreed that the highlighted area covers the nerve region on a scale from 0-10. Moderators then presented results to the raters and proceeded to the next image if and only if they both determined that no major disagreements existed between both parties. In the case of a major disagreement, both moderators initiated a discussion period after having presented the different results. A chance was then presented to the raters to revise their previous answers to the questions. Descriptive statistics were generated from these results. The methods used were Cohen's kappa, binary logistic regressions, and ROC curves.

Medical experience survey

Upon meeting with each individual labeler, a brief survey of ten questions was presented gauging their experience in the field of anesthesia. The survey questions can be found in Appendix I. As with the previous experiments, a regression analysis was run on this data paired with each labeler's average Yes/No score.

Results

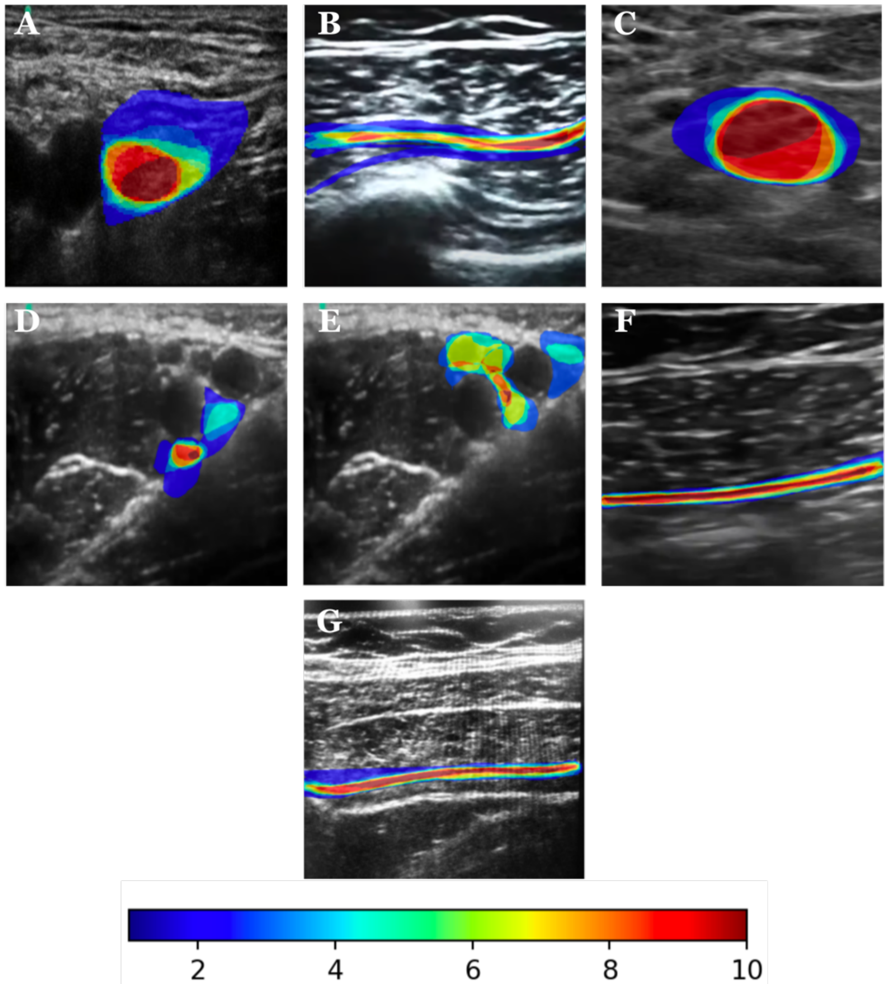

Figure 1. Consensus heat maps depicting the range of areas commonly agreed upon among the 10 anesthesiologists. The selected images shown to each physician are as follows: (A) Femoral nerve, (B) Pecs I, (C) Sciatic nerve at the popliteal fossa, (D) Radial nerve, (E) Median and ulnar.

As seen in Figure 1, which illustrates the levels of agreement among the labelers, the Axillary block presented higher than normal variability. All descriptive statistics in each experiment were performed with and without Axillary-related results, effectively forming four groups: All physicians, all nerves (APAN); all physicians, no Axillary (APnA); experts only, all nerves (EAN); experts only, no Axillary (EnA). For each of these groups, ten tables such as Supplemental Data Table 1 were recorded for each physician with the 1/0 encodings representing their Yes/No answers to the corresponding labels. These answers were then appended to the condensed pairwise metric sheet and a binary logistic regression was computed for each group.

An ICC was computed for APAN and APnA, receiving a score of 0.65 and 0.08, respectively. Where the former indicates an overall moderate agreement according to Koo and Li's ICC scale, APnA's ICC analysis breaks down due to the failure to incorporate chance where there is high agreement.[16] The average agreement for APnA without including chance is 85.87%. EAN received a kappa score of 0.76 which represents substantial agreement according to Landis and Koch's kappa scale.[17] However, EnA breaks down when using descriptive statistics that incorporate chance as both experts agreed on all images (Yes) except in one instance (Yes and No). Out of a total of 100 answers (50 images, 2 experts), 98 were the same; therefore, we observed a 98% agreement for EnA.

The full regression analysis performed on the binary yes/no survey data vs. each pair-wise computed metric can be found in Table 1. Additional analysis on more machine-learning-specific metrics such as the MCC and Threat score can be found in the Supplemental Data Table 3. The ROC curves for each metric and group can be seen below in Figure 2. Furthermore, a summarized and complete table of the ROC analysis can be found in Table 2. MCC and Threat score ROC results can be found in Supplemental Data Table 4.

Yes/No – Score Binary Logistic Regression Analysis

Metric	Group	Regression coefficient	OR	95% Confidence Interval	p-value
Accuracy	APAN	-0.0769	0.9259	(0.84,1.02)	0.1252
	APnA	0.1826	1.2004	(1.07,1.34)	**0.0012**
	EAN	-0.6006	0.5485	(0.37,0.82)	**0.0033**
	EnA	0.1217	1.1294	(0.55,2.30)	0.7377
Sensitivity	APAN	0.0209	1.0211	(1.02,1.03)	<<**<0.05**
	APnA	0.0041	1.0041	(0.99,1.02)	0.5571
	EAN	0.0089	1.0089	(0.99,1.02)	0.1866
	EnA	0.0804	1.0837	(0.94,1.25)	0.2683
Specificity	APAN	0.0522	1.0536	(0.96,1.16)	0.2759
	APnA	0.2372	1.2677	(1.14,1.42)	<<**<0.05**
	EAN	-0.3272	0.7209	(0.46,1.14)	0.1621
	EnA	-0.0694	0.9329	(0.21,4.23)	0.9283
PPV	APAN	0.0268	1.0272	(1.02,1.03)	<<**<0.05**
	APnA	0.0280	1.0284	(1.02,1.04)	<<**<0.05**
	EAN	0.0167	1.0168	(1.00,1.03)	**0.0192**
	EnA	0.0519	1.0533	(0.97,1.14)	0.1901
NPV	APAN	-0.2579	0.7726	(0.66,0.91)	**0.0016**
	APnA	-0.0878	0.9159	(0.75,1.12)	0.3822
	EAN	-0.5509	0.5764	(0.36,0.93)	**0.0242**
	EnA	0.1014	1.1067	(0.60,2.04)	0.7460
Dice score	APAN	0.0285	1.0289	(1.02,1.04)	<<**<0.05**
	APnA	0.0249	1.0252	(1.01,1.04)	**0.0007**
	EAN	0.0149	1.015	(1.00,1.03)	**0.0486**
	EnA	0.0663	1.0685	(0.97,1.18)	0.1961

Table 1. Binary logistic regression results for each metric and the physicians' yes/no responses within each group.

Yes/No – Score ROC Analysis

Metric	Group	threshold	AUC	sensitivity	1-specificity
Accuracy	APAN	95.92	0.45	0.54	0.51
	APnA	96.14	0.64	0.76	0.53
	EAN	98.54	0.27	0.00	0.00
	EnA	95.79	0.72	0.72	0.00
Sensitivity	APAN	73.48	0.67	0.83	0.48
	APnA	82.84	0.47	0.90	0.76
	EAN	51.81	0.60	0.65	0.45
	EnA	77.25	0.92	0.92	0.00
Specificity	APAN	98.89	0.56	0.53	0.36
	APnA	95.60	0.71	0.61	0.18
	EAN	96.35	0.40	0.12	0.05
	EnA	99.92	0.56	0.56	0.00
PPV	APAN	53.23	0.73	0.84	0.44
	APnA	72.83	0.68	0.52	0.21
	EAN	59.90	0.67	0.62	0.32
	EnA	79.16	0.90	0.90	0.00
NPV	APAN	99.17	0.43	0.00	0.00
	APnA	98.94	0.45	0.85	0.79
	EAN	98.64	0.34	0.04	0.00
	EnA	98.92	0.56	0.56	0.00
Dice score	APAN	34.44	0.72	0.73	0.33
	APnA	81.35	0.61	0.74	0.47
	EAN	64.67	0.65	0.62	0.32
	EnA	59.63	0.90	0.90	0.00

Table 2. Summary table of ROC analysis for each metric vs. Yes/no results.

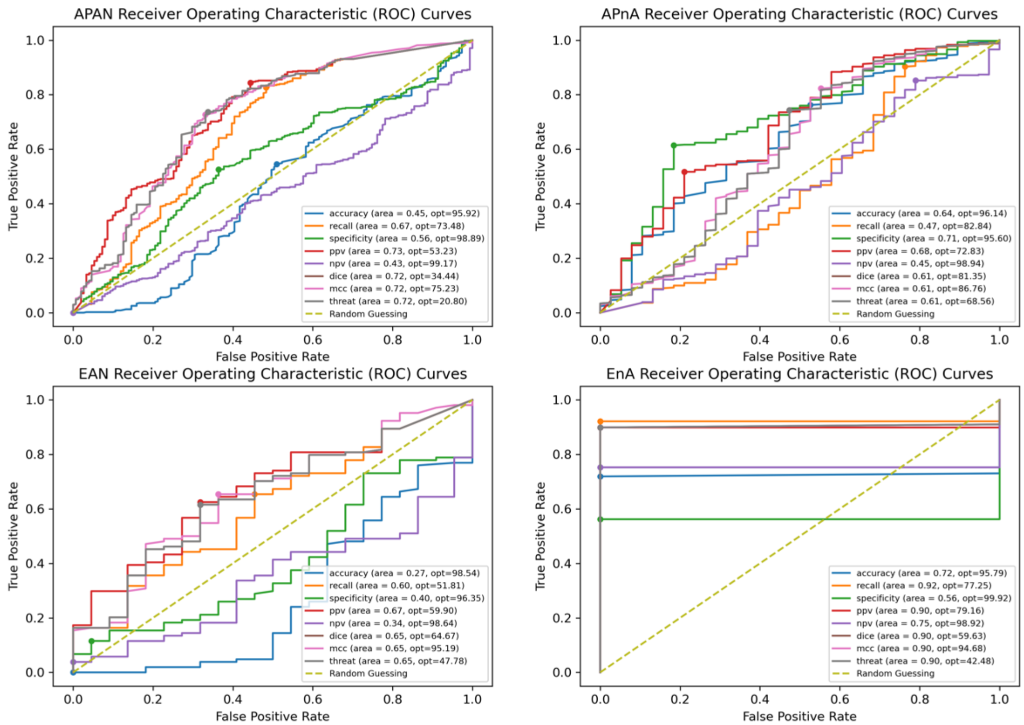

Figure 2. ROC curves for each metric and yes/no results within each group.

The logistic regression between the yes/no answers (encoded 0 for no and 1 for yes) and the 0-10 scores produced a dependent variable coefficient of 2.021 (p=0.0003). Moreover, the ROC curves seen in Figure 3 show areas under the curve (AUC) of 0.99 each and an optimal 0-10 score of 7/10 for both experts.

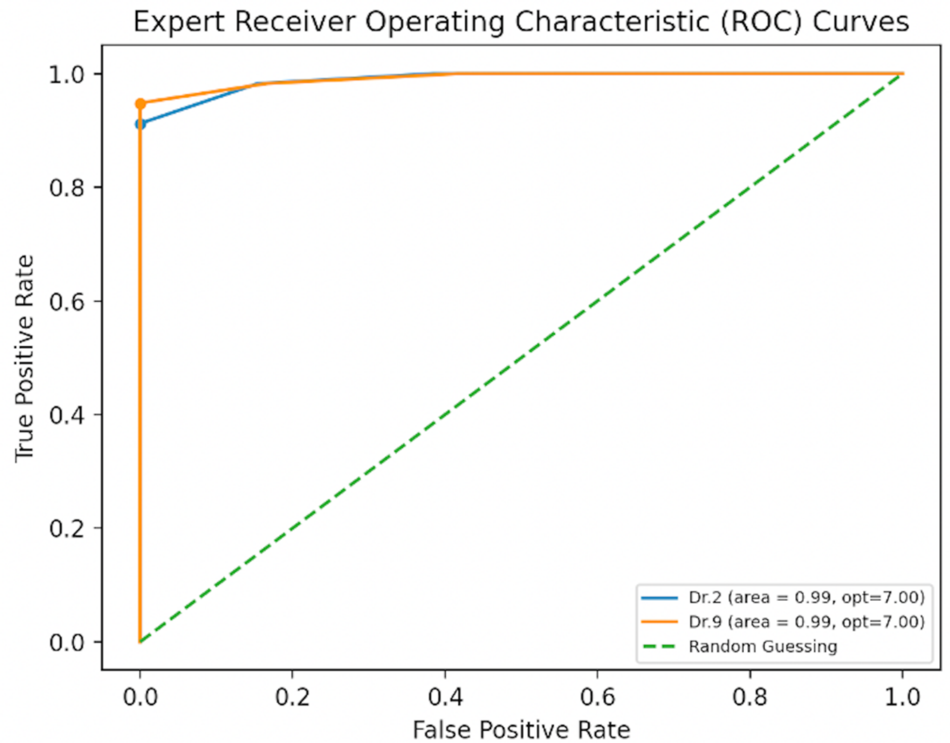

Figure 3. ROC curves of both experts' 0-10 evaluations vs their Yes/No responses.

Lastly, the regression analysis performed on the physicians' experience in the field versus their average Yes/No score can be seen in Table 3.

Regression Analysis on Anesthesiologist Expertise

Independent Variable (vs. Average Yes/No score)	Regression coefficient	p-value
Completed a fellowship	-3.68	0.576
Years of experience	**-0.571**	**0.045**
No. blocks per month	-0.0237	0.698

Table 3. Regression analysis on all physicians' experience vs. average Yes/No score.

Discussion

Deep learning medical image segmentation tasks are best paired with Dice score evaluation schemes when determining model performance, as they provide an optimal indicator of segmentation quality when compared to similar metrics.[18] For this reason, the medical imaging literature primarily utilizes the Dice metric as a standard for validation and performance interpretation.[19] The issue with this current approach stems from a lack of transparency and inability to quantify or visualize what a high Dice score transfers to in practice. Research tends to emphasize high scores without critically translating model performance to real-world clinical significance. As a result, numerous clinical research teams have reported difficulties in utilizing models outside of research settings.[20,21] Here, we attempted to explore US nerve target zone interpretation by equating model performance to that of experts in the field, with the intention being an objectively defined Dice score threshold that relates to clinical effectiveness.

For every group except EnA, the group demonstrating near perfect agreement, there are statistically significant relationships between the Dice metric and the labelers' Yes/No answers – inter- and intra-labeler agreement. A positive regression coefficient of 0.0285 and odds ratio (OR) of 1.0289 ($p < 0.05$, 95% CI = [1.02,1.04]) observed in APAN indicated a positive and directly proportional relationship between the Dice score and the likelihood of obtaining a Yes answer from experts. Every incremental Dice coefficient increase of 1% was found to be associated with an approximately 3% increase in the likelihood that a physician will agree the highlighted region is safe for needle placement. Similarly, the ROC analysis on the Dice metric showed that the APAN group received the highest AUC of 72%, yielding an optimal Dice score of 34.44% ($p < 0.05$). These results objectively quantify that a minimum Dice score of 34% is required to minimise the false positive rate, defined as one minus specificity, while maximizing the true positive rate of the

Yes/No predictions. These observations lead to a surprising finding, which appears to show that a previously considered low Dice score can be otherwise linked to a high and statistically significant agreement amongst labelers. This in turn might be an indication into how researchers can treat future clinical relevance when it comes to automatic nerve segmentation in US images. The inter- and intra-expert analysis on the physician's Yes/No and 0-10 ratings of the images produced an ROC curve with an AUC of 0.99, indicating a near-perfect classifier, with an optimal threshold of 0.7. These results indicate a statistically significant relationship between the experts' 0-10 rating of all the cohort's highlighted regions and their Yes/No answers. A score of 7/10 was found to therefore be optimal for having the highest odds of obtaining a Yes answer.

Regarding physician experience, it is of particular interest to see how an inversely proportional and statistically significant relationship exists between the years of experience in medicine and a labeler's average Yes/No score. In our case, a physician's average Yes/No score decreases by a factor of 0.571 for each increasing year of experience ($p=0.045$). Residents and newly attending anesthesiologists may be more likely to have been recently exposed to a wider range of sonoanatomy cases, as well as having recently undergone multiple examinations similar to the tasks given in this study. These factors could account for the apparent difference in scores by experienced and newly attending physicians.

Figure 1 clearly demonstrates a high degree of agreement between labellers for the planar blocks (I.e., TAP, Pecs I, and Posterior Rectus Sheath) or ones featuring large regions of interest (I.e., FN and SN). A limitation that can be associated with the vast differences in labelled regions for the nerves in the Axillary BP might be that the orientation and placement of the US probe resulting in the shown scan was not provided, potentially confusing the labellers on where anatomical landmarks might be positioned. In a study that examined the shifting sonoanotomy of the Axillary

BP, 69 healthy volunteers agreed to be scanned by US medical professionals.[22] The aim of this study was to investigate expert accuracy in locating the median, ulnar, and radial nerves in the US feeds at three distinct levels: between the Pectoralis Major and Bicep (A), between levels A and C (B), and at the largest bicep circumference (C).[22] Results have shown that at level A, the chances of correctly locating the nerves range from 30%-59% with all nerves being visible in all of the images. In between 9%-13% of cases at level B, at least one nerve has disappeared from the image, increasing up to 30%-80% at level C.[22] It is therefore evident that acquiring the complete sonoanatomy of the Axillary BP, particularly that of location and orientation, is necessary to perform a block; successful visualization being highly dependent on probe placement. This issue can be mitigated in future works by providing US probe orientation information along with each image, allowing experts to better distinguish block target zones. An additional limitation in this study was that only one randomly chosen image in each nerve dataset containing 10 images each was shown to the anesthesiologists. A more thorough statistical analysis would be achieved had we performed the Yes/No survey on all of the compiled images. This would not only increase the number of questions from 70 to 700 but would have also increased the time taken to complete the survey by a factor of 10. With the average time taken to complete the survey being 5 minutes, requesting upwards of 50 minutes was not a viable option for this preliminary study.

Previous research on expert comparisons of ultrasound examinations lack objective metrics, which in turn results in the inability of future researchers to compare novel AI models' segmentation performance against that of a human equivalent labelled test set. For instance, the current standard in the field of radiology involves determining differences between sonographers and expert radiologists, but these comparisons fail to report on important statistical findings beyond simple discrepancy rates between both groups .[23-25] Dawkins and colleagues found a 15.5% discrepancy between sonographers and radiologists' interpretation of biliary findings in upper right quadrant US images but was not statistically significant.[24] In another recent paper, the authors chose to define their own diagnostic scoring criteria based on categorical differences in agreement, reporting a minor discrepancy rate of 2.8% between radiologists and radiographers in sonographic findings.[25] Although useful in determining differences between human labellers, this methodology provides no quantifiable insight on what it means to be an expert. In the future we recommend researchers attempt to determine objective and statistically relevant thresholds for clinical relevancy.

In this study, our goal was to establish an objective and quantifiable link between inter- and intra-labeler agreement on various nerve locations in ultrasound images. This was accomplished through a two-step process in which ten physicians' labels were compared to one another using common comparison metrics. Furthermore, each physician rated their own label as well as the labels of each of their colleagues by selecting yes or no to the question of whether the label is appropriate for performing a nerve block. By performing statistical analysis to obtain a statistically significant relationship between an objective metric and a subjective rating (Yes/No), this study serves as a benchmark for future studies involving nerve segmentation that wish to justify their models' performance.

Appendix I

Expert questionnaire:

1. Please enter your name
2. Have you completed a fellowship?
3. If yes to the above, what was the focus of your fellowship? (e.g., regional anesthesia, cardiac anesthesia)
4. How many years have you been practicing after your residency/fellowship?
5. On average, how many TAP blocks do you perform on a monthly basis?
6. On average, how many PECS (I and/or II) blocks do you perform on a monthly basis?
7. On average, how many Rectus Sheath blocks do you perform on a monthly basis?
8. On average, how many Femoral Nerve blocks do you perform on a monthly basis?
9. On average, how many Popliteal Fossa Sciatic Nerve blocks do you perform on a monthly basis?
10. On average, how many Axillary Brachial Plexus blocks do you perform on a monthly basis?

Supplemental Data

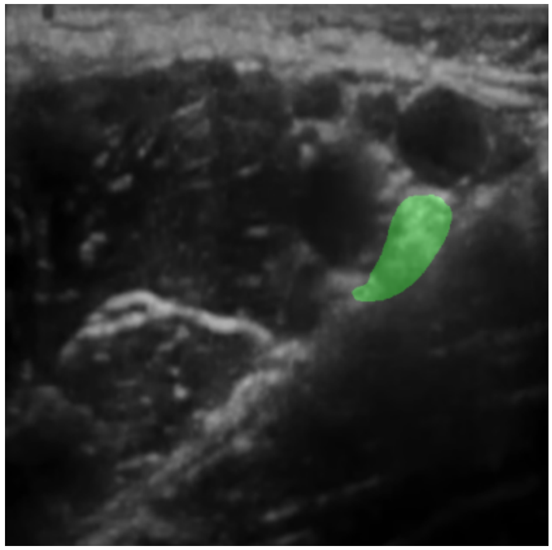

Figure 1. An anesthesiologist's label of the Radial Nerve in an ultrasound image of the Axillary Brachial Plexus.

Sample Yes/No Labels from Dr. 1

	Femoral Nerve Block	TAP Block	Popliteal Block	Rectus Sheath Block	Axillary BP Block		
Labelers /region	**FN**	**TAP**	**SN**	**Posterior sheath**	**MN_UN**	**RN**	**Pecs I**
Dr. 1	0	1	0	1	0	1	1
Dr. 2	1	1	0	1	0	0	0
Dr. 3	1	1	1	1	0	0	1
Dr. 4	0	1	0	1	0	0	1
Dr. 5	0	0	0	1	0	0	1
Dr. 6	1	1	0	1	1	1	1
Dr. 7	1	1	0	1	0	0	1
Dr. 8	0	1	1	0	0	1	0
Dr. 9	1	1	1	1	0	0	1
Dr. 10	0	1	0	1	0	0	1

Table 1. A table containing Dr. 1's binary encoded yes/no answers for each of their colleagues' respective labels as well as their own.

Overall Inter- and Intra- Rater Scores

	Femoral Nerve Block	TAP Block	Popliteal Block	Rectus Sheath Block	Axillary BP Block		Pecs Block	Score (/70)
Labelers	**fn**	**tap**	**sn**	**posterior sheath**	**mn_un**	**rn**	**pecs1**	**Score**
Dr. 1	9	10	9	10	6	9	10	90.00
Dr. 2	8	10	8	10	4	2	9	72.86
Dr. 3	10	10	9	10	6	1	9	78.57
Dr. 4	8	10	9	9	1	1	10	68.57
Dr. 5	8	7	9	10	2	1	9	65.71
Dr. 6	10	10	9	10	6	10	10	92.86
Dr. 7	9	10	9	10	2	2	9	72.86
Dr. 8	6	10	10	8	3	10	9	80.00
Dr. 9	9	10	10	10	4	2	4	70.00
Dr. 10	8	10	9	9	3	2	10	72.86
% agreement/region	85	97	91	96	37	40	89	

Table 2. The sum of every raters' binary encoded yes/no answers for each of their colleagues' respective labels as well as their own over the seven nerve regions.

Yes/No – Score Binary Logistic Regression Analysis

Metric	Group	Regression coefficient	OR	95% Confidence Interval	p-value
MCC	APAN	2.7513	15.663	(8.42,29.13)	**<<<0.05**
	APnA	2.4671	11.7887	(2.9,47.83)	**0.0006**
	EAN	1.3730	3.9470	(0.94,16.54)	0.0604
	EnA	5.9339	377.628	(0.06,25082)	0.1826
Threat score	APAN	0.0335	1.0341	(1.03,1.04)	**<<<0.05**
	APnA	0.0238	1.0241	(1.01,1.04)	**0.004**
	EAN	0.0193	1.0195	(1.00,1.04)	**0.0348**
	EnA	0.0986	1.1036	(0.93,1.31)	0.2588

Table 3. Binary logistic regression results performed on the MCC and Threat score metrics versus the respective labeler pair's Yes/No rating.

Yes/No – Score ROC Analysis

Metric	Group	threshold	AUC	sensitivity	1-specificity
MCC	APAN	75.23	0.72	0.73	0.33
	APnA	86.76	0.61	0.82	0.55
	EAN	95.19	0.65	0.65	0.36
	EnA	94.68	0.90	0.90	0.00
Threat score	APAN	20.80	0.72	0.74	0.34
	APnA	68.56	0.61	0.74	0.47
	EAN	47.78	0.65	0.62	0.32
	EnA	42.48	0.90	0.90	0.00

Table 4. ROC results for the MCC and Threat score metrics when compared to the labelers' Yes/No ratings.

Contributions

Noam Suissa: This author conceived and designed the study, searched the literature, extracted the data, interpreted the data, performed the statistical analysis, and drafted the manuscript.

Sean D Jeffries: This author searched the literature, extracted the data, interpreted the data, drafted the manuscript.

Jose Luis Ramirez-GarciaLuna**:** This author aided in statistical analysis and critically revised the manuscript.

Joshua Morse: This author aided in editing of the manuscript and study design.

Thomas M. Hemmerling: The corresponding author who conceived and designed the study, interpreted the data and critically revised the manuscript.

All authors read and approved the final manuscript.

References

1. Rahimzadeh P, Faiz SH. Ultrasound a new paradigm in regional anesthesia and pain management. Anesth Pain Med. 2013;3(2):228-229. doi:10.5812/aapm.13363
2. Fujiwara Y, Ito H, Shibata Y, Komatsu T. *Masui*. 2008;57(5):543-548.
3. McCartney CJ, Lin L, Shastri U. Evidence basis for the use of ultrasound for upper-extremity blocks. Reg Anesth Pain Med. 2010;35(2 Suppl):S10-S15. doi:10.1097/AAP.0b013e3181d25675
4. Henderson M, Dolan J. Challenges, solutions, and advances in ultrasound-guided regional anaesthesia. BJA Education. 05 2016;16(11):374-380. doi:10.1093/bjaed/mkw026
5. Pesapane, F., Codari, M. & Sardanelli, F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp 2, 35 (2018). https://doi.org/10.1186/s41747-018-0061-6
6. Huang C, Zhou Y, Tan W, et al. Applying deep learning in recognizing the femoral nerve block region on ultrasound images. Annals of Translational Medicine. 2019 Sep;7(18):453. DOI: 10.21037/atm.2019.08.61. PMID: 31700889; PMCID: PMC6803209
7. Gungor I, Gunaydin B, Oktar SO, et al. A real-time anatomy ıdentification via tool based on artificial ıntelligence for ultrasound-guided peripheral nerve block procedures: an accuracy study. J Anesth. 2021;35(4):591-594. doi:10.1007/s00540-021-02947-3
8. Bowness, J., Varsou, O., Turbitt, L., Burkett-St Laurent, D. (2021). Identifying anatomical structures on ultrasound: assistive artificial intelligence in ultrasound-guided regional anesthesia. Clinical Anatomy, 34(5), 802– 809. https://doi.org/10.1002/ca.23742

9. Bowness JS, El-Boghdadly K, Woodworth G, et alExploring the utility of assistive artificial intelligence for ultrasound scanning in regional anesthesiaRegional Anesthesia & Pain Medicine 2022;47:375-379
10. Wang Y, Zhu B, Kong L, et al. Brachial Plexus Nerve Trunk Segmentation Using Deep Learning: A Comparative Study with Doctors' Manual Segmentation. arXiv [eessIV]. Published online 2022. http://arxiv.org/abs/2205.08143
11. Github. femoral_nerve_block_computer_vision. Available from https://github.com/gscfwid/femoral_nerve_block_computer_vision (Accessed 22 November 2022)
12. Hadjerci O, Hafiane A, Conte D, Makris P, Vieyres P, Delbos A. Computer-aided detection system for nerve identification using ultrasound images: A comparative study. Informatics in Medicine Unlocked. 2016;3:29-43. doi:10.1016/j.imu.2016.06.003
13. Baby M, Jereesh AS. Automatic nerve segmentation of ultrasound images. In: 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA). Vol 1. ; 2017:107-112. doi:10.1109/ICECA.2017.8203654
14. Hicks, S.A., Strümke, I., Thambawita, V. et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep 12, 5979 (2022). https://doi.org/10.1038/s41598-022-09954-8
15. Github. Swipe-Labeler. Available from https://github.com/spaceml-org/Swipe-Labeler (Accessed 10 December 2022)
16. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research [published correction appears in J Chiropr Med.

2017 Dec;16(4):346]. J Chiropr Med. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174.
18. Eelbode T, Bertels J, Berman M, et al. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. IEEE Transactions on Medical Imaging. 2020;39(11):3679-3690. doi:10.1109/TMI.2020.3002417
19. Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res Notes*. 2022;15(1):210. Published 2022 Jun 20. doi:10.1186/s13104-022-06096-y
20. Altaf F, Islam SMS, Akhtar N, Janjua NK. Going Deep in Medical Image Analysis: Concepts, Methods, Challenges, and Future Directions. IEEE Access. 2019;7:99540-99572. doi:10.1109/ACCESS.2019.2929365
21. Chen, H., and Sung, J. J. Y. (2021) Potentials of AI in medical image analysis in Gastroenterology and Hepatology. Journal of Gastroenterology and Hepatology, 36: 31–38. https://doi.org/10.1111/jgh.15327.
22. Retzl G, Kapral S, Greher M, Mauritz W. Ultrasonographic findings of the axillary part of the brachial plexus. Anesthesia & Analgesia. 2001;92(5):1271-1275. doi:10.1097/00000539-200105000-00037
23. Lo RH, Chan PP, Chan LP, Wilde CC, Pant R. Routine abdominal and pelvic ultrasound examinations: an audit comparing radiographers and radiologists. Ann Acad Med Singap. 2003;32(1):126-128.

24. Dawkins A, George N, Ganesh H, et al. Radiologist and Sonographer Interpretation Discrepancies for Biliary Sonographic Findings: Our Experience. Ultrasound Q. 2017;33(4):261-264. doi:10.1097/RUQ.0000000000000280
25. Williams I, Baird M, Schneider M. Comparison between radiographers with sonography education working in remote Australia and radiologists' interpretation of ultrasound examinations. J Med Radiat Sci. 2022;69(3):293-298. doi:10.1002/jmrs.576

Literature Review

This section presents a detailed review of the relevant literature in the fields of ultrasound-guided regional nerve blocks, their associated risks and benefits, and AI techniques applied to enhance the outcomes of these procedures. From a total of 166 articles, 72 accepted articles have been analyzed and summarized in this section. Note that, on few occasions, editorials and other minor publications have been accepted due to their important citations. Furthermore, the table below gives a brief overview of how this review is broken down.

Category	Description	Number of articles
AI in medicine	Medical applications, trends, and recommendations regarding artificial intelligence in medicine	9
AI in anesthesiology	Trends and limitations of AI in anesthesiology	8
Education in anesthesia	Trends in anesthesia education as well as new approaches for better technique in regional anesthesia	7
Medical Image segmentation	Novel machine and deep learning models and trends that accurately highlight target regions in medical images	23
Overview of various blocks	Reviews and overviews of relevant nerve block procedures in different scenarios	9
Risks and benefits of UGPNB	Adverse and positive outcomes of specific nerve block procedures without the use of artificial intelligence	16

AI in medicine

Predicting adverse outcomes of cardiac surgery with the application of artificial neural networks (Peng et al., 2008)

In this article, the authors use and artificial neural network (ANN) to predict the mortality and morbidity of patients post-cardiac surgery. Furthermore, they compare the model's results to a logistic regression model and a Parsonnet score. The ANN outperformed the other models and metrics in its accuracy for predicting major morbidity, and its area under the curve (AUC) of the receiver operating characteristic for both in-hospital mortality and major morbidity.

Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine (Pesapane et al., 2018)

This article discusses the emerging technological applications for artificial intelligence in the medical field, specifically medical imaging. The authors define key terms such as machine and deep learning and show how the number of published articles on the topic has increased by 600% from 2007 to 2017. However, despite these technologies performing better at detecting important pathologies that might be invisible to the naked than experienced physicians, a total replacement from doctors to AI models will not likely happen. Instead, they will work hand in hand, with AI providing radiologists with a helping hand and alleviating their workload so that they may be more visible to a larger volume of patients.

Machine learning in medicine: a practical introduction (Sidey-Gibbons et al., 2019)

Sidey-Gibbons and colleagues demonstrate the use of popular machine learning techniques to accurately provide cancer diagnoses from cell nuclei in breast mass samples. While the article focuses solely on machine learning algorithms as opposed to the purely deep learning ones used in this thesis, the authors present important information as a guide on how to develop these algorithms which is universally similar in all subfields of AI. Furthermore, the paper presents crucial statistical tools that can be used to evaluate model performance.

Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine (Ahmed et al., 2020)

The authors discuss how precision medicine is a promising advancement in medical care that can improve the traditional symptom-based approach by allowing earlier interventions and personalized treatments through advanced diagnostics. To achieve this, they argue that it is necessary to analyze comprehensive patient information, which can help identify biological indicators that signal shifts in health. Technological advancements have made it easier to utilize healthcare information in clinical decision-making, but integrating disparate data sources and addressing ethical and social issues related to privacy and protection of healthcare data is crucial. Furthermore, machine learning platforms can support clinicians by efficiently analyzing and managing clinical data to optimize decision-making. The authors close off by stating that the use of artificial intelligence has the potential to lead to significant improvements in personalized and population medicine at lower costs, but academic solutions are necessary to pave the way for this new era of discovery in healthcare.

A short guide for medical professionals in the era of artificial intelligence (Meskó et al., 2020)

In this article, the authors present an objective view of the current state of artificial intelligence (AI) in medicine. Their primary purpose is to educate physicians on the basic definitions of AI, Machine/Deep Learning, their applications in various sub-specialties in medicine, how to evaluate news and studies about AI, as well as the future of AI in medicine and the potential obstacles that might lie ahead.

Recommendations for Reporting Machine Learning Analyses in Clinical Research (Stevens et al., 2020)

In this article, the authors discusses the challenges that come with interpreting ML models and their results in a clinical setting. On one hand, medical experts and peer reviewers might over or underestimate ML models' performances, and on the other, ML experts who do not have medical experience may present their results in a fashion that is too hard to assess by physicians (Stevens et al., 2020). To address this issue, the authors propose a methodology for presenting ML techniques and results that are easily and readily interpretable by medical professionals.

Machine learning in medicine: It has arrived, let's embrace it (Pappada, 2021)

In this article, the author discusses the growth of artificial intelligence and machine learning in multiple patient care settings and demonstrates how these technologies have direct positive impacts on patient care outcome predictions, clinical decision support, and therapeutic setpoint prediction. Furthermore, the article focuses on the use of machine learning to predict one-year patient survival post-orthotopic heart transplantation procedures. The use cases for such a complex operation include the improvement of clinical decision-making, patient counselling, and organ allocation. The author concludes that AI and machine learning are here to make considerable impacts in patient care in the near future.

Deep Learning for Medical Anomaly Detection A Survey (Fernando et al., 2021)

In this article, the authors explore the advantages and limitations of deep learning methods in medical anomaly detection found in a lack of structure across the numerous published studies on the matter. The authors focus on providing a comprehensive analysis of popular deep-learning techniques for medical anomaly detection by presenting a coherent and systematic review of the current methods and by comparing and contrasting their architectural difference and training algorithms. Lastly, the article outlines important limitations of current deep-learning medical anomaly detection algorithms and suggests further research to address them.

On evaluation metrics for medical applications of artificial intelligence (Hicks et al., 2022)

The authors present a way to standardize the reporting of an ML model's performance with the use of eight metrics. These metrics, namely, accuracy, recall, specificity, precision (positive and negative predictive value), F1 or Dice score, Matthews correlation coefficient, and the threat score, provide some sort of insight into how well a model's prediction matches a ground truth label. Furthermore, these metrics are all calculated by four values: true positives, false positives, true negatives, and false negatives. With simple operations performed on a combination of these values, the eight metrics arise. Lastly, the authors argue that including all of these eight metrics in an ML study can not only increase the comprehensibility of the article but also be used to compare to other ones that lack some of the metrics.

AI in anesthesiology

Assessment of a simple artificial neural network for predicting residual neuromuscular block (Laffey et al., 2003)

This article presents the use of an artificial neural network to accurately predict whether a patient is experiencing Postoperative residual curarization (PORC). Of 40 recruited patients, the neural network achieved a negative predictive value of 0.93 vs 0.35 provided by the anesthetic care providers. The authors conclude by stating that the use of a neural network was very practical for estimating the likelihood of PORC at the time of tracheal extubation. Furthermore, they state that the application of neural networks to other predictive problems in anesthesia may be beneficial.

Neural nets and prediction of the recovery rate from neuromuscular block (Santanen et al., 2003)

Similar to the article presented by Laffey et al. (2003), this article presents how multiple neural networks can be used to predict a rough estimate of the remaining recovery time following a neuromuscular block during general anesthesia. The models would look at four variables that are known to affect the block such as multiple minimum alveolar concentration, end-tidal CO2 concentration, and peripheral and central temperature (Santanen et al., 2003). The authors conclude that the trained neural networks better predict the recovery time than the average-based method used in the study.

Use of Machine Learning Theory to Predict the Need for Femoral Nerve Block Following ACL Repair (Tighe et al., 2011)

The authors present machine learning strategies to predict the requirement of postoperative femoral nerve blocks (FNB) following anterior cruciate ligament (ACL) reconstruction. While it is common to perform a FNB preoperatively in order to reduce postoperative pain following the surgical reconstruction, factors causing the need to perform a postoperative FNB are currently unclear. Therefore, multiple ML models using different statistical strategies were employed to train on data gathered by 349 patients who have undergone ACL reconstruction. Overall, the authors report that all ML models outperformed traditional statistical methodologies in this particular task and can provide improved predictive capabilities (Tighe et al., 2011).

Anesthesiology, automation, and artificial intelligence (Alexander et al., 2018)

In this editorial, the authors argue that algorithm-based automation in anesthesiology has remained a difficult task. While their use has led to improved patient outcome and the requirement of less anesthetic than usual, full autonomy is virtually impossible. This is due in part to the nature of algorithms being built from a set of specific rules. When applied to a procedure that has so many potential outcomes, one may not be caught and handled leading to potential risks for the patient. However, with the advent of AI and machine learning, intelligent algorithms capable of learning from past experience can be utilized to overcome this challenge especially in the realm of clinical decision making.

Artificial Intelligence and Machine Learning in Anesthesiology (Connor, 2019)

In this review article, Connor argues how commercial applications that integrate artificial intelligence are sometimes prone to produce to tolerable errors. However, in anesthesia, there is no room for errors, and constant attention and response to feedback is crucial to patient outcomes. Furthermore, in this review, Connor presents concepts on how to link artificial intelligence to anesthesiology with relevant clinical questions.

Artificial Intelligence in Anesthesiology: Current Techniques, Clinical Applications, and Limitations (Hashimoto et al., 2020)

In this review, the authors present the potential for artificial intelligence and its use in anesthesiology. Furthermore, the authors summarize six themes where AI can play a role in this medical field, namely, depth of anesthesia monitoring, control of anesthesia, event and risk prediction, ultrasound (US) guidance, pain management, and operating room logistics (Hashimoto et al., 2020). They go on to explain how this can be achieved with machine learning, deep learning, and traditional AI techniques. Lastly, after having presented some of AI's important limitations in regards to these application, the authors conclude that machine learning techniques applied to anesthesia can potentially improve patient and physician care perioperatively.

Artificial intelligence for image interpretation in ultrasound-guided regional anaesthesia (Bowness et al., 2020)

In this editorial, the authors discuss the challenges that anesthesiologists face when interpreting sonoanatomy in patients undergoing ultrasound-guided regional nerve block procedures and how machine and deep learning can provide a solution to this issue. Moreover, the authors present important studies on how machine learning applications can be applied to the field of anesthesia perioperatively. Lastly, one such tool for automatically highlighting important anatomical structures viewed in an ultrasound for the adductor canal block is presented and shown to help minimize unwanted trauma from the needle. The authors conclude by stating that anesthesiologists should embrace deep learning solutions as they can positively impact patient care and outcomes.

Exploring the utility of assistive artificial intelligence for ultrasound scanning in regional anesthesia (Bowness et al., 2022)

In this brief technical report, the authors explore the utility of a deep learning tool, ScanNav, capable of highlighting nerve regions in ultrasound feeds in real time to assist anesthesiologists in performing nerve block procedures. Thirty anesthesiologists, 15 expert and 15 non-expert, performed 240 ultrasound scans across nine peripheral nerve block regions (Bowness et al., 2022). Furthermore, each participant was asked to complete a questionnaire on the relevant impact this tool can bring to the procedure. Non-experts show more positive feedback than experts ($p=0.0001$) (Bowness et al., 2022). Additionally, both groups show that this software has a potential for being used as a training tool. Lastly, less than five percent of experts reported a potential risk increase for approximately 10 scans in the set (Bowness et al., 2022). The authors conclude that ScanNav can aid in generalizing standard care in ultrasound-guided regional anesthesia.

Education in anesthesia

The American Society of Regional Anesthesia and Pain Medicine and the European Society Of Regional Anaesthesia and Pain Therapy Joint Committee recommendations for education and training in ultrasound-guided regional anesthesia (Sites et al., 2009)

This special article contains work from a joint committee from ASRA and the European Society of Regional Anesthesia and Pain Therapy whereby recommendations on how to establish guidelines for the main disciplines in regional anesthesia such as teaching and the scope of practice (Sites et al., 2009). More specifically, the document outlines four topics: a list of ten commonly performed tasks when conducting Ultrasound-Guided Regional Anesthesia (UGRA), crucial set of skills needed to perform these procedures, and training pathways for both postgraduate anesthesiologists and residents. The document concludes by stating that it is the joint committee's recommendation that the requirements for granting UGRA rights to anesthesiologists is up to the individual institution's discretion.

Ultrasound and its evolution in perioperative regional anesthesia and analgesia (Mariano et al., 2014)

In this article, the authors provide a history of regional anesthesia and UGRA, their clinical applications, the evidence basis for their use, as well as future trends in the respective fields (Mariano et al., 2014).

Development and Validation of an Assessment of Regional Anesthesia Ultrasound Interpretation Skills (Woodworth et al., 2015)

In this study to develop an assessment tool for evaluating ultrasound interpretation skills for regional anesthesia, a 50-question assessment was created based on the inputs from residents, academic faculty, community anesthesiologists, and expert video recordings. After pilot testing and final administration to 90 participants, a 47-question subset was found to be reliable and valid, with a significant correlation between expected and predicted item difficulty. Moreover, test scores linearly increased with higher levels of formal anesthesia training, regional anesthesia training, the number of ultrasound-guided blocks performed per year, and self-assessment of regional anesthesia skill. This test could serve as an effective tool for competency milestone assessment in anesthesiology training.

Challenges, solutions, and advances in ultrasound-guided regional anaesthesia (Henderson et al., 2016)

In this article, UGRA is described to offer significant advantages for delivering analgesia to patients compared to traditional landmark-based techniques previously used in the field (Henderson et al., 2016). However, despite these benefits, UGRA relies on visual continuous image interpretation which is crucial for successfully and safely implementing a block. Furthermore, ultrasound feeds are not always perfect and devoid of artificial artifacts, ultimately introducing many challenges that anesthesiologists might face when interpreting the sonoanatomy for correctly placing the needle. The article covers a wide variety of challenges that can be encountered in US feeds. For example, acoustic artifacts are produced by the US probe's sound waves interacting in a particular way with the various types of tissue resulting in image artifacts that may hinder the process of sonoanatomy interpretation. Moreover, the paper discusses other challenges such as anatomical pitfall errors, optical illusions, oedema, obesity, air present in the tissue, and many more to look out for when performing UGRA. Lastly, the authors conclude by presenting promising technologies and potential advancements that may aid in addressing these challenges altogether.

A pragmatic approach to evaluating new techniques in regional anesthesia and acute pain medicine (Mudumbai et al., 2018)

In this very short Perspective, the authors present a framework for deciding whether to improve upon existing techniques in regional anesthesia or create new ones. Each technique should be evaluated through these four factors: increase in access, enhancements in efficiency, decreases disparity, and improves outcomes. By assessing a current or proposed technique through these four categories, the authors argue that this framework will help clinicians better make decisions on whether they should adopt or improve new and/or existing techniques.

Future directions in regional anaesthesia: not just for the cognoscenti (Turbitt et al., 2020)

In this editorial, the authors present three key components for improving regional anesthesia and have its practices implemented to the greatest possible patient population (Turbitt et al., 2020). The first step, achieving widespread implementation of common nerve blocks, consists of standardizing the adoption of skills required to perform small, basic blocks that are usually easy to learn. These procedures include the Interscalene and Axillary Brachial Plexus (BP) blocks, Femoral nerve, Adductor canal, Popliteal Sciatic nerve, Erector spinea, and Rectus sheath blocks.

The next step is to ensure that competency in performing these blocks is maintained by adjusting the curriculum and improving practical skills in the field. The final step is to implement this acquired knowledge into clinical pathways. To achieve this, multiple cultural and systematic barriers must be overcome.

Defining an Ultrasound-guided Regional Anesthesia Curriculum for Emergency Medicine (Tucker et al., 2021)

In this study aimed at identifying components of an UGRA curriculum for emergency medicine (EM) physicians, an expert panel voted on potential curriculum elements using a modified Delphi process. Although UGRA offers numerous benefits, many EM trainees lack focused education in this area. The expert panel reached a high level of agreement for 65 background knowledge elements and ten UGRA techniques. The resulting curriculum can serve as a foundation for developing comprehensive UGRA education for both residents and independent providers in EM.

Medical Image Segmentation

Nerve Localization by Machine Learning Framework with New Feature Selection Algorithm (Hadjerci et al., 2015)

In this study, the authors propose a nerve localization framework with a new feature selection algorithm for UGRA. The method, based on statistical approaches and learning models, aims to assist anesthetists by automating nerve detection in ultrasound images, a challenging task due to noise and artifacts. The results indicate that the proposed method accurately and efficiently identifies nerve zones, outperforming existing techniques with an accuracy of 82% on one dataset and 61% on another untrained dataset.

U-net: Convolutional networks for biomedical image segmentation (Ronneberger et al., 2015)

In this article, the authors present a novel architecture for deep learning-based segmentation of biomedical images. The authors introduce the U-Net, a convolutional neural network (CNN) designed specifically to address the challenges of segmenting biomedical images, such as limited training data and the need for precise boundary localization. The U-Net architecture consists of a contracting path for feature extraction and an expanding path for precise localization, connected by skip connections. The authors demonstrate the effectiveness of U-Net by applying it to various datasets, including electron microscopic images of neural structures and histological sections of the kidney. The results show that U-Net outperforms existing methods and achieves high accuracy, making it a valuable tool for biomedical image segmentation tasks.

Computer-aided detection system for nerve identification using ultrasound images: A comparative study (Hadjerci et al., 2016)

In their paper on UGRA, the authors address the challenge of nerve detection and segmentation in ultrasound images. They propose an efficient framework for nerve detection and segmentation, while also reviewing and evaluating the performance of existing methods in the literature. The proposed system comprises four main stages: (1) de-speckling filter, (2) feature extraction, (3) feature selection, and (4) classification and segmentation. The authors conducted a comparative study on each stage to measure its impact on the overall system. Using sonographic videos from 19 volunteer patients, they assessed the effect of training set size and evaluated consistency through a cross-validation technique. The proposed framework achieved high scores (80% on average of 1900 tested images), demonstrating its validity and potential usefulness in UGRA applications.

Assistive system based on nerve detection and needle navigation in ultrasound images for regional anesthesia (Hadjerci et al., 2016)

In this study, the authors present the first fully automatic system for detecting regions of interest and generating needle trajectories in UGRA. The system addresses two critical steps in UGRA: anatomical structure recognition and needle steering towards the target region. The proposed system consists of two stages. The first stage involves the automatic localization and segmentation of nerves and arteries in ultrasound images using a machine learning algorithm with a multi-model classification process and an active contour. The second stage involves the development of a path planning algorithm to obtain the optimal needle insertion trajectory based on the results of the first stage. The effectiveness of the proposed system was evaluated through experiments on individual modules of the detection framework and by comparing the overall framework to existing methods. Two datasets, acquired at different times, were used to assess the robustness of the proposed method. Experimental results demonstrate the robustness and feasibility of the proposed assistive system in UGRA practice, potentially improving its safety and generalizing it to medical facilities with limited practitioner expertise.

Automatic Nerve Segmentation Of Ultrasound Images (Baby et al., 2017)

In this conference paper, the authors propose a method to segment US images of the BP nerve bundle. Their process starts by passing the training set (n=5640) through a de-speckling filter to reduce background speckle noise and then on to training using the popular U-Net architecture. Training was also performed on a traditional Support Vector Machine (SVM) for performance comparison. On 5508 test images, the mean Dice score for the U-Net and SVM were 0.71 and 0.64, respectively.

Automatic Segmentation and Probe Guidance for Real-Time Assistance of Ultrasound-Guided Femoral Nerve Blocks (Smistad et al., 2017)

The authors of this study propose a system to assist inexperienced physicians in performing ultrasound-guided femoral nerve blocks. The system guides the user in moving the ultrasound probe to investigate the region of interest and reach the target site for needle insertion. It also provides automatic real-time segmentation of the femoral artery, the femoral nerve, the fascia lata and fascia iliaca, aiding in the interpretation of the 2D ultrasound images and surrounding anatomy in 3D. The system was evaluated on 24 ultrasound acquisitions from six subjects and the results were compared to those of an expert anaesthesiologist. The average target distance was 8.5 mm with a standard deviation of 2.5 mm and the mean absolute differences of the femoral nerve and fascia segmentations were about 1-3 mm.

Improved U-Net Model for Nerve Segmentation (Zhao et al., 2017)

In this study, the authors present a method for automatic segmentation of medical images using CNNs, capitalizing on the advancements in computer vision. The proposed network architecture, based on the U-Net model, employs inception modules and batch normalization instead of standard convolutional layers, effectively reducing the number of parameters and accelerating training without sacrificing accuracy. The authors also substitute the binary cross entropy loss function with the Dice coefficient. Their proposed model scored an average Dice score of 0.653 with a model size of 5M parameters whereas the U-Net scored 0.658 with 31M parameters.

Segmentation of nerve on ultrasound images using deep adversarial network (Liu et al., 2018)

In this study, the authors develop a deep adversarial neural network to address the challenges associated with segmenting the BP nerve on US images. The authors established a segmentation network based on a variation of the VGG network. They then incorporated a discriminator network to ensure the anatomical dependencies which evaluates the quality of segmentation. Lastly, elastic deformations are introduced to the dataset to mimic anatomic variations across multiple patient profiles. A mean intersection over union (mIOU) score of 73.29% was achieved for the authors' model including deep adversarial networks.

Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks (Smistad et al., 2018)

In this study, the authors utilize a deep convolutional neural network to identify key structures like nerves and blood vessels in ultrasound images collected during axillary nerve block procedures. They compile a dataset of 49 subjects to train and evaluate the neural network. Different image augmentations, including rotation, elastic deformation, shadows, and horizontal flipping, are assessed. The authors perform cross-validation to evaluate the neural network, and find that blood vessels were detected most easily with a precision and recall above 0.8. Among the nerves, the median and ulnar nerves are detected most effectively with F-scores of 0.73 and 0.62 respectively, while the radial nerve proves to be most challenging to detect with an F-score of 0.39. The authors note that image augmentations improved the F-score by as much as 0.13, with the combination of all augmentations providing the best results. However, they acknowledge that the precision and recall values are still not optimal and suggest that a larger dataset, combined with anatomical and temporal models, might be required to enhance accuracy.

Deep convolutional neural network for segmentation of knee joint anatomy (Zhou et al., 2018)

In this study, the authors present a novel method for knee joint tissue segmentation, integrating a CNN, 3D fully connected conditional random field (CRF), and 3D simplex deformable modeling. This strategy delivers high-resolution pixel-wise tissue classification, ensuring the contextual voxel relationships are maintained and the joint structure shape is preserved. The evaluation using 3D fast spin-echo MR image datasets yielded impressive results: high mean Dice coefficients above 0.9 for four tissue types and between 0.7 and 0.9 for eight others, demonstrating strong accuracy and robustness of the method. This approach suggests promising potential for enhancing efficiency and accuracy in knee joint tissue segmentation in musculoskeletal imaging.

Deep visual nerve tracking in ultrasound images (Alkhatib et al., 2019)

In this study, the researchers provide a comparative analysis of thirteen recent deep-learning trackers on different types of nerves, assessing their accuracy, consistency, time complexity, and adaptability to diverse nerve situations, such as loss of shape information or tissue disappearance. Testing of these trackers are performed on a median and sciatic nerve US dataset consisting of 10,337 still images captured from 42 adult patients. The findings indicate that these deep-learning trackers offer robust performance across different kinds of nerves, affirming their potential in facilitating UGRA procedures.

Medical image segmentation algorithm based on feedback mechanism convolutional neural network (Feng-Ping and Zhi-Wen, 2019)

In this study, the authors address the limitations of traditional image segmentation methods and standard CNNs in the field of medical image segmentation. They propose a new algorithm that is inspired by the feedback mechanism of the human visual cortex. Two new algorithms based on the greedy strategy are proposed to solve the feedback optimization problem. The authors then present a medical image segmentation algorithm that leverages this feedback mechanism within the CNN. This involves learning and extracting deep image features through unlabeled image block sample training to construct feedback mechanism convolutional neural network models. These models are then used to classify pixel block samples in the medical image to be segmented, with further optimization through threshold segmentation and morphological methods. The proposed method demonstrates high segmentation accuracy and adaptability for various medical images.

Applying deep learning in recognizing the femoral nerve block region on ultrasound images (Huang et al., 2019)

In this study, the authors developed a method to identify the femoral nerve block region in ultrasound images, aimed primarily at less experienced operators. They collected and annotated a dataset of ultrasound images showing the femoral nerve block. They used the U-net framework to train the model, which segmented the region of interest in the images. Model performance was evaluated based on Intersection over Union (IoU) and accuracy metrics. The median IoU results for the training set, development set, and test set were 0.722, 0.653, and 0.644 respectively, while the segmentation accuracy of the test set was 83.9%. Moreover, 10-fold cross-validation resulted in a median IoU of 0.656 and accuracy between 82.1% and 90.7%. The authors conclude that the trained model using U-net demonstrated satisfactory performance in femoral-nerve region segmentation and may have potential for clinical application.

High performance neural network inference, streaming, and visualization of medical images using FAST (Smistad et al., 2019)

The authors of this paper highlight the use of inference engines (IEs) for executing deep convolutional neural networks for medical image analysis, noting that existing IEs like Intel's OpenVINO, NVIDIA's TensorRT, and Google's TensorFlow work with specific processors and have distinct APIs. They propose methods to extend the FAST framework, which is an open-source, high-performance tool for medical imaging, to work with any IE via a common programming interface. This would simplify deploying and testing neural networks on different processors. The proposed approach is evaluated on three tasks: real-time ultrasound image segmentation, CT volume segmentation, and patch-wise classification of whole slide microscopy images. They find significant performance variations depending on the IE and processor combination, with differences in processing times ranging from half a second to 24 seconds for ultrasound frames, 2 to 53 seconds for volume processing, and 81 seconds to nearly 16 minutes for processing whole slide microscopy images.

NAS-Unet: Neural architecture search for medical image segmentation (Weng et al., 2019)

In this paper, the authors extend neural architecture search (NAS) to medical image segmentation, inspired by the success of U-net and its variants in this field. They create three types of primitive operation sets for the search space and use these to find two cell architectures, DownSC and UpSC. These are used in NAS-Unet, a U-shaped network for semantic segmentation. The DownSC and UpSC architectures are updated concurrently using a differential architecture strategy. The authors test their method on Promise12, Chaos, and ultrasound nerve datasets collected via MRI, CT, and ultrasound respectively. Their NAS-Unet, trained on PASCAL VOC2012, showed superior performance and had significantly fewer parameters than U-net and one of its variants when tested on the mentioned medical image datasets.

Artificial intelligence in detection and segmentation of internal auditory canal and its nerves using deep learning techniques (Jeevakala et al., 2020)

This study introduces an automated method for detecting and segmenting the internal auditory canal (IAC) and its associated nerves using a Mask R-CNN approach combined with U-net. The RESNET50 model, the backbone of Mask R-CNN, localizes the IAC, while U-net is used to learn the features and segment the IAC and its nerves. The method was tested on clinical datasets from 50 patients, both adults and children. Evaluation metrics included IoU for IAC localization, and Dice similarity coefficient for segmentation. Results showed the method had an impressive performance with RESNET50 and RESNET101 achieving a mean IoU of 0.79 and 0.74 respectively. In terms of segmentation, the method scored higher Dice similarity coefficient than region growing and Particle Swarm Optimization (PSO) methods, at 96%. The results suggest the proposed AI tool can aid radiologists by providing accurate localization and segmentation of the IAC and its nerves.

Self-co-attention neural network for anatomy segmentation in whole breast ultrasound (Lei et al., 2020)

This study presents an automatic breast anatomy segmentation method for automated whole breast ultrasound (AWBUS) images to support image interpretation and breast density estimation. The researchers tackle issues such as low image quality and ill-defined boundaries by developing a new deep learning encoder-decoder segmentation method based on a self-co-attention mechanism. This mechanism includes a spatial and channel attention module (SC) in the ResNeXt (Res-SC) block and a non-local context block (NCB) for learning high-level context. The decoder path employs a weighted up-sampling block (WUB) for better class-specific up-sampling, while a co-attention mechanism improves segmentation coherence between consecutive slices. Extensive experiments and comparisons with other leading deep learning segmentation methods validate the effectiveness of the proposed method for breast anatomy segmentation in AWBUS images.

Identifying anatomical structures on ultrasound: assistive artificial intelligence in ultrasound-guided regional anesthesia (Bowness et al., 2021)

This study examines the usefulness of an AI system in aiding the identification of anatomical structures during ultrasound-guided regional anesthesia. Due to common difficulties among anesthesiologists in recognizing these structures, an AI system was tested for its potential benefits. Three regional anesthesia experts reviewed 40 ultrasound scans, comparing unmodified and AI-highlighted videos side-by-side. They rated the system's overall performance, assessed the utility of highlighting for identifying specific structures, and considered its aid in confirming correct ultrasound views for less experienced practitioners. The AI system demonstrated helpfulness in identifying specific anatomical structures in nearly all cases (99.7%) and in confirming the correct ultrasound view in 99.3% of scans. Despite a need for further evaluation, these findings illustrate the potential of such AI technology to enhance clinical practice and revitalize the field of clinical anatomy.

A self-spatial adaptive weighting based u-net for image segmentation (Cho et al., 2021)

The study proposes a novel spatially adaptive weighting scheme for medical image segmentation, aiming to enhance the performance of U-Net-based architectures. The scheme utilizes various convolutional frameworks like VGG, ResNet, and Bottleneck ResNet structures, and substitutes the up-convolutional layer with a bilinear up-sampling method. Evaluation of this method on three different medical imaging datasets showed notable improvements in segmentation performance. Specifically, the network with the proposed self-spatial adaptive weighting block based on the ResNet framework yielded the highest IoU and Dice scores among tested methods. In particular, for the Nerve dataset, the combination of the proposed block and the ResNet framework led to an increase of 3.01% in IoU and 2.89% in Dice scores, thus demonstrating the significant potential of this new approach for image segmentation tasks.

A real-time anatomy identification via tool based on artificial intelligence for ultrasound-guided peripheral nerve block procedures: an accuracy study (Gungor et al., 2021)

The study assessed the precision of a real-time AI-based anatomy identification tool created to support UGRA image interpretation. A variety of nerve blocks were carried out using the software on 40 healthy subjects (20 women and 20 men) by anesthesiology students. The ultrasound images were saved and afterwards examined by professional validators once the software confirmed 100% scan success of anatomical landmarks for each block. When trainees had 100% scan success, the validators' accuracy ratings were consistent. Except for transversus abdominis plane (TAP) blocks, which exhibited an inverse connection with age and BMI, the scores did not significantly change according on participant demographics. These findings imply that AI can assist anesthesiologists in performing UGPNB by accurately interpreting anatomical structures in real-time sonography.

Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment (Marzola et al., 2021)

In this study, the authors present a deep learning solution to segment cross sectional areas (CSA) in transversus musculoskeletal US images while also providing a quantitative grayscale analysis in the images. The dataset is composed of 3917 images from 1283 subjects and labeled by experts in the field. Bland-Altman plots, grayscale analysis, and correlation analysis were used to compare the automatic segmentation predictions to those of the experts in the test set. A precision of 0.88 ± 0.12 and a recall of 0.92 ± 0.09 were achieved in the test set but were slightly lower for abnormal muscles. Additionally, intra-class and Pearson's correlation coefficients demonstrated strong agreement in the analysis. The CSA segmentation model demonstrated satisfactory performance and provided grayscale z-score information to the likes of manual operators.

Lesion segmentation in breast ultrasound images using the optimized marked watershed method (Shen et al., 2021)

In order to enhance the segmentation of lesions in breast ultrasound (BUS) images, the study proposes the Adaptive Morphological Snake Marked Watershed (AMSMW) technique. The approach outperformed both the RDAU-NET deep learning model and conventional segmentation techniques in the study's tests on two datasets. On dataset A (n=500 private BUS images), it achieved a 96.25% accuracy rate, 78.4% Dice similarity coefficient, and a 65.34% Jaccard index. Additionally, it attained a sensitivity of 88.79%, a Dice similarity coefficient of 86.25%, and an accuracy of 97.96% on dataset B (n=205 open source BUS images). These findings point to the method's potential to improve breast cancer screening, particularly in remote places without access to qualified radiologists.

Brachial Plexus Nerve Trunk Recognition From Ultrasound Images: A Comparative Study of Deep Learning Models (Tian et al., 2022)

This study aims to compare the performance of twelve deep learning models for the segmentation of the brachial plexus. The dataset was composed of 340 BP images annotated by three anesthesiologists. Of the twelve models evaluated, the U-Net architecture achieved the highest mean IoU of 68.5%, however, is limited to process 15 frames per second due to the model's large size. On the other hand, the LinkNet architecture came in second place with a mean IoU score of 66.27% and was capable of processing 142 frames per second.

Overview of various blocks

Peripheral Nerve Blocks Improve Analgesia After Total Knee Replacement Surgery (Allen et al., 1998)

This study examined the effectiveness of a femoral nerve block (FNB) for total knee replacement (TKR) surgeries which are commonly associated with severe postoperative pain. Furthermore, on top of studying the analgesic effects of FNBs, the authors wanted to see that of a combined femoral-sciatic nerve block. A control group receiving a "sham" block was assessed. For eight hours after transfer to the hospital ward, patients receiving nerve blocks reported less pain ($p < 0.05$). Moreover, morphine use dropped by 50% up until the second day postop for the groups receiving nerve blocks ($p < 0.02$). No statistically significant difference in the analgesic effects between FNBs and femoral-sciatic blocks were observed.

Ultrasound imaging in anesthesia: an overview of vascular access and peripheral nerve blocks (Sandhu, 2007)

In this article, the author describes the benefits and limitations of US imaging for nerve blocks and provides general descriptions of the different transducer types and approaches to enter different anatomical regions with a needle. Moreover, visual guides and methods of previous studies from the following blocks are given: interscalene, supraclavicular, infraclavicular, axillary, femoral, sciatic nerve, saphenous, obturator, Ilioinguinal and hypogastric, cervical, occipital, and neuraxial blocks.

Ultrasound guidance for deep peripheral nerve blocks: a brief review (Wadhwa et al., 2011)

This article reviews the efficacy of the use of ultrasound in peripheral nerve blocks compared to the traditional nerve stimulation technique. This study focuses on infraclavicular, lumbar plexus, and sciatic nerve blocks. Furthermore, the findings indicate that transitioning to US-guided blocks alone or in tandem with nerve stimulation can be beneficial in certain scenarios, however, can pose challenges for deeper blocks or novice practitioners who are unable to properly interpret sonoanatomy.

Four quadrant transversus abdominis plane block and continuous transversus abdominis plane analgesia: a 3-year prospective audit in 124 patients (Niraj et al., 2015)

This prospective study examined the effectiveness of a new technique to provide continuous analgesia in the TAP region for patients undergoing emergency or elective abdominal surgery (Niraj et al., 2015). The study's cohort comprised of 124 adult patients scheduled to undergo elective or emergency abdominal surgery. As for the results, 70% of patients received incisions within the dermatomal limit of the block. Furthermore, the analgesic failure rate was 10% with 39% of patients reporting no failure within the first 48 hours; only 57% experienced less than five episodes. The authors conclude that the four quadrant TAP block is an effective technique for offering perioperative analgesia especially when the surgery is performed near the dermatomal limit.

Update on ultrasound for truncal blocks: a review of the evidence (Abrahams et al., 2016)

This article updates a previous systematic review conducted in 2010 by the authors on the evidence behind several truncal blocks performed using ultrasound. This updates brings about three new US-guided block studies and a new systematic review was conducted to provide new recommendations on the blocks. These blocks include the paravertebral, intercostal, transversus abdominis plane, rectus sheath, ilioinguinal/iliohypogastric, as well as the Pecs, quadratus lumborum, and transversalis fascia blocks. Thanks to these new studies, the authors conclude that our understanding of the anatomy pertinent to these blocks may improve as well as evaluating patient-related risks and outcomes.

Abdominal wall blocks in adults (Børglum et al., 2016)

In this review article, the authors investigate the efficacy of abdominal wall blocks using US guidance. They argue that recent findings show that UGRA in abdominal wall blocks is the gold standard in adults with the TAP block being the most commonly performed. Finally, the authors conclude that abdominal wall blocks are assigned a grade A or B evidence based on the guidelines by the US Agency for Healthcare Policy and Research.

Ultrasound-Guided Regional Anaesthesia: Visualising the Nerve and Needle (Bowness and Taylor, 2020)

In this review article, the authors present an overview of UGRA along with its benefits and challenges. Moreover, the authors provide in-depth descriptions and visual guides of the sonoanatomy for the following blocks: Interscalene block, Supraclavicular block, Infraclavicular block, Axillary brachial plexus block, Forearm blocks, Femoral nerve block, Popliteal sciatic nerve block, and ankle block. Strategies on how to properly place the US probe for optimal nerve and needle visualization is also provided.

Ultrasound-guided transgluteal sciatic nerve analgesia for refractory back pain in the ED: A case series (Goldsmith et al., 2020)

In this case series, the authors present an US-guided transgluteal sciatic nerve block (TGSNB) procedure performed on three patients experiencing sciatic radicular back pain in the emergency department (ED) setting. Since there is a limited pain management regiment in the ED, this study aids in providing a useful and cost-effective use case for UGRA procedures in this setting. All three patients experienced no adverse outcomes and reported very good pain relief (Goldsmith et al., 2020).

New composite scale for evaluating peripheral nerve block quality in upper limb orthopaedics surgery (Almasi et al., 2021)

In this article, the authors argue that a robust tool for evaluating the quality of US-guided nerve blocks does not exist. Therefore, the study presents a method for evaluating US-guided interscalene-supraclavicular blocks and axillary-supraclavicular blocks on 93 patients. Sensory, motor, coping, and postoperative pain (SMCP) metrics were recorded for each patient as well as the quality of the anesthesia graded by an anesthesiologist (QAGA). Results showed that no significant difference in QAGA was observed for both block groups. Furthermore, 97.8% of patients were in the Excellent and Good categories with SMCP whereas 86% were with QAGA ($p<0.001$) (Almasi et al., 2021).

Risks and benefits of UGPNB

Transversus Abdominis Plane Block: How Safe is it? (Jankovic et al., 2008)

In this short letter to the editor, the authors note that TAP blocks are considered to be safe and effective with minimal complications. Several risks, however, do include organ needle puncture and accidental anesthetic intraperitoneal injections which can cause organ damage. Some suggestions are then offered to minimize the risks of this procedure such as decreasing the volume of anesthetic used and using a blunt-edged needle to avoid unnecessary tissue puncture.

Ultrasound-guided peripheral nerve blocks: What are the benefits? (Koscielniak-Nielsen, 2008)

In this review study, the author examined the growing use of ultrasound in anaesthesiology for regional blocks and found that ultrasound guidance offers significant benefits in clinical practice. Moreover, after scanning MEDLINE and EMBASE databases, it was established that when peripheral nerves are well imaged by ultrasound, there's no added advantage to using nerve stimulation, however, obtaining satisfactory images can sometimes be challenging. The use of ultrasound significantly reduced block performance time, the number of needle passes, and block onset time. The occurrence of paraesthesia during the procedure was also less frequent, although there were contradictory results regarding the incidence of accidental vascular punctures. Post-operative neuropraxia incidence remained unchanged. In pediatric patients, the duration of the block was found to be longer, but this was not the case in adults. Furthermore, the study revealed that ultrasound guidance allowed for a reduction in the dose of local anaesthetic used in blocks.

Limitations and Technical Considerations of Ultrasound-Guided Peripheral Nerve Blocks: Edema and Subcutaneous Air (Saranteas et al., 2008)

This case report presents two trauma patients for whom the benefits of ultrasound-guided peripheral nerve blocks were limited due to edema and subcutaneous air. The first patient, affected by tissue edema and obesity with a body mass index of 35, and the second patient, dealing with subcutaneous emphysema, experienced difficulties with 2-dimensional ultrasound imaging despite the use of advanced equipment and techniques. Consequently, the use of neurostimulation technique alone or in combination with ultrasound imaging was necessary to successfully perform the nerve block.

Complications of peripheral nerve blocks (Jeng et al., 2010)

This article draws upon the known complications resulting from peripheral nerve blocks. However rare these complications might be, the consequences on both the patient and anesthesiologist must be considered. Furthermore, the review focuses on complications that can arise from nerve blocks, continuous peripheral nerve catheter techniques, and local anesthetic toxicity (Jeng et al., 2010).

Liver trauma secondary to ultrasound-guided transversus abdominis plane block (Lancaster & Chadwick, 2010)

In this correspondence by the editor, a case is described where a 61 year old male patient was found to have liver damage after undergoing a TAP block to treat a strangulated inguinal hernia.

Transversus abdominis plane block: a note of caution! (Walker, 2010)

In this short correspondence, the editor references a cadaveric study on the TAP block which finds that the femoral nerve lies parallel to the deeper parts of the transversus abdominis plane. Therefore, as found by the study, uncareful placement of the needle and minimal injectate could lead to a femoral nerve palsy.

Probable Local Anesthetic Systemic Toxicity in a Postpartum Patient with Acute Fatty Liver of Pregnancy After a Transversus Abdominis Plane Block (Naidu & Richebe, 2013)

This case report presents the complications associated with peripheral nerve blocks, specially TAP blocks, for a 25 year old patient with acute fatty liver of pregnancy.

Cardiac Arrest from Local Anesthetic Toxicity After a Field Block and Transversus Abdominis Plane Block: A Consequence of Miscommunication Between the Anesthesiologist and Surgeon (Scherrer et al., 2013)

Another case report is presented where a 25 year old female patient who has undergone laparoscopic gynecologic surgery under general anesthesia experienced a seizure followed by ventricular arrhythmia. This occurred after a bilateral TAP block was administered by an anesthesiologist due to the patient experiencing severe postoperative pain.

Feasibility and analgesic efficacy of the transversus abdominis plane block after single-port laparoscopy in patients having bariatric surgery (Wassef et al., 2013)

This study examines the analgesic efficacy of US-guided TAP blocks in morbidly obese patients undergoing single-port sleeve laparoscopic gastrectomy (SPSG). The authors found that patients who received TAP blocks experienced less severe pain 6 to 12 hours post-surgery compared to the intravenous control group. Nevertheless, the total opioid consumption 24 hours post-surgery was similar in both groups. The authors conclude that US-guided TAP blocks are effective in morbidly obese patient populations and provide immediate pain relief following SPSG.

Demonstrating the Benefits of Transversus Abdominis Plane Blocks on Patient Outcomes in Laparoscopic Colorectal Surgery: Review of 200 Consecutive Cases (Keller et al., 2014)

In this study, the long-term effects of TAP blocks and enhanced recovery protocols on 200 patients who have undergone colorectal resections is evaluated. When performed together, the authors observed a median length of stay of 2 days, where a majority of 77% discharged by postop day 3. Moreover, the complication rate was 12% and readmission rate was 6.5%. The authors conclude that the combination of TAP blocks and a standardized enhanced recovery protocol improved colorectal resection results, reduced length of stay and maintained lower readmission rates.

Convulsions in 2 Patients After Bilateral Ultrasound-Guided Transversus Abdominis Plane Blocks for Cesarean Analgesia (Weiss et al., 2014)

This report presents seizure cases in two patients who received bilateral TAP blocks following a cesarian section operation; a procedure that has been associated with no major complications. Both incidents required resuscitation and bag-mask interventions, respectively, and lipid emulsions before fully recovering. The authors provide a note of caution for anesthesiologists who offer TAP block in post-cesarian delivery patients and warn of potential anesthetic toxicity.

Use of Transversus Abdominis Plane (TAP) Blocks for Pain Management in Elderly Surgical Patients (Sammons & Ritchey, 2015)

This article discusses the effectiveness of the TAP block for postoperative analgesia in elderly patients. Common comorbidities are often tied to this patient population and minimizing postoperative complications is important. The study found that administering TAP blocks in this patient category can mitigate risks such as pneumonia, urinary retention, and falls.

Clinical effectiveness of transversus abdominis plane (TAP) blocks for pain relief after caesarean section: a meta-analysis (Champaneria et al., 2016)

This review of 20 randomized control trials studied the effectiveness of TAP blocks for managing postoperative pain in patients who have undergone cesarean sections. The findings suggest that TAP blocks significantly reduced patients both at rest and on movement compared to placebo and intrathecal morphine. However, the significant difference is lost when comparing TAP to intrathecal morphine or co-administration.

Efficacy of transversus abdominis plane block with liposomal bupivacaine during open abdominal wall reconstruction (Fayezizadeh et al., 2016)

This authors studied the analgesic effects of TAP blocks performed on patients undergoing abdominal wall reconstruction while having liposomal bupivacaine (LB) injections. The study found that patients who received both TAP and LB experienced less postoperative pain and a shorter length of stay at the hospital compared to the LB-only control group. The authors conclude that intraoperative TAP blocks paired with LB can significantly improve perioperative patient outcomes.

The use of ultrasound guided combined peripheral nerve blocks in a high-risk patient: A case report (Kavakli et al., 2021)

This study explores the use of multiple peripheral nerve blocks, namely, a femoral, anterior sciatic, and lateral femoral cutaneous nerve block for above-knee amputations in high risk patients.

Safety and effectiveness of ultrasound guided peripheral nerve blocks: Audit at tertiary care hospital (Salam et al., 2021)

The aim of this retrospective study was to analyze the analgesic efficacy of peripheral nerve blocks administered to patients in the Aga Khan University Hospital in Karachi from 2015-2017. Numerous variables were monitored in the 299 patients such as pain scores, complications, and need for additional analgesia. Overall, the results showed that UGRA is safe and effective in providing postoperative analgesia. Furthermore, 70% of patients in the study reported fully effective analgesia on movement 12 hours after surgery.

Discussion & Conclusion

Two studies focusing on improving the field of ultrasound-guided regional anesthesia have been conducted. The first, a deep learning solution to automatically highlight the transversus abdominis plane region in ultrasound images resulted in a 73% global Dice score against a test set of ten images each labeled by ten anesthesiologists (Suissa et al., 2023, In Press). Secondly, another study attempted to establish a clinically-relevant comparison metric threshold so that ML researchers who attempt US nerve segmentation can base their results on (Suissa et al., 2023, Pending approval). The study found that a statistically significant Dice metric threshold of 34% would determine whether or not a pixel-wise comparison would be deemed clinically acceptable for the specific application of UGPNB.

The first study was considered to be a proof of concept to test our methods for the specific task of highlighting a nerve in a set of ultrasound images. Its success proved that these methods should be applied to a wider range of nerve blocks, each of which feature unique and challenging sonoanatomy. Furthermore, by being equipped with a clinically relevant threshold for a set of seven nerve regions, future studies can begin to evaluate the segmentation efficacy of DL models on these nerves as well as new ones.

The second study, currently pending approval from the British Journal of Anesthesia, provided key insight on how we should treat DL nerve segmentation results from published studies. Whereas the traditional way of evaluating a model's performance is to compare it with past studies' results for a specific task (e.g. nerve segmentation), there have never been, to our knowledge, studies that use a clinically-backed evaluation benchmark. By finding a statistically significant Dice score threshold of 34%, the first study's 73% global Dice score is deemed to be clinically relevant. One particular limitation that might not have been mentioned in the paper is that the evaluated thresholds are task-specific and can only be used for segmentation evaluation for the seven nerve regions presented in the study. Additional research using similar methods would have to be conducted to obtain new thresholds for a wider range of nerve regions/blocks. Additionally, these methods can also be applied to new modes of computer vision deep learning tasks such as pathology detection in MRI or CT scans.

Finally, a comprehensive literature review covering 72 peer-reviewed research articles, case reports, and editorials on a myriad of subjects revolving around anesthesiology and deep learning were presented. The first two sections of the review provided a thorough overview of the state of artificial intelligence in the last decade in medicine as a whole and anesthesiology, respectively. The third section provided a glimpse at the education curriculum for anesthesiologists wanting to perform nerve blocks, from challenges to overcome in the field to new approaches in existing nerve block procedures. In the Medical Image Segmentation section of the review, over 20 scholarly articles presenting novel techniques and deep learning architectures were summarized. These studies demonstrated advancements in nerve segmentation in ultrasound feeds as well as detecting other tissues and blood vessels hoping to assist anesthesiologists better perform nerve blocks. The final two sections in the review covered the proper methods to recognize sonoanatomy for performing effective blocks as well as highlight the risks and benefits that anesthesiologists should be aware of concerning these procedures. Overall, the articles summarized in this review provide a thorough overview for anesthesiologists and machine learning researchers of the state of artificial intelligence applied to the automatic detection of nerves in US images.

The studies presented in this thesis demonstrated a method to gauge clinical relevance in US nerve segmentation methods as well as using it to validate our custom deep learning software capable of highlighting the TAP region in various images. With future enhancements to the research, additional improvements can be made to strengthen the statistical power of the results with the first step being to increase the evaluated test cases in each experiment. Furthermore, new state-of-the-art DL architectures should be explored to optimize the final models' predictive scores and effectively bring researchers closer to achieving a clinically-accepted decision support system for anesthesiologists.

References

Abrahams M, Derby R, Horn J-L. Update on ultrasound for truncal blocks: a review of the evidence. Regional Anesthesia & Pain Medicine. 2016 2016;41(2):275. doi:doi:10.1097/AAP.0000000000000372

Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database : the journal of biological databases and curation. 2020 2020;2020:baaa010. doi:doi:10.1093/database/baaa010

Alexander JC, Joshi GP. Anesthesiology, automation, and artificial intelligence. Baylor University Medical Center Proceedings. 2018 2018;31(1):117-119. doi:doi:10.1080/08998280.2017.1391036

Alkhatib M, Hafiane A, Vieyres P, Delbos A. Deep visual nerve tracking in ultrasound images. Computerized Medical Imaging and Graphics. 2019 2019;76doi:doi:10.1016/j.compmedimag.2019.05.007

Allen HW, Liu SS, Ware PD, Nairn CS, Owens BD. Peripheral Nerve Blocks Improve Analgesia After Total Knee Replacement Surgery. Anesthesia & Analgesia. 1998 1998;87(1)doi:doi:

Almasi R, Rezman B, Kovacs E, et al. New composite scale for evaluating peripheral nerve block quality in upper limb orthopaedics surgery. Injury. 2021 2021;52:S78-S82. doi:doi:10.1016/j.injury.2020.02.048

Baby M, Jereesh AS. Automatic nerve segmentation of ultrasound images. 2017:107-112.

Børglum J, Gögenür I, Bendtsen TF. Abdominal wall blocks in adults. Current Opinion in Anesthesiology. 2016 2016;29(5)doi:doi:

Bowness J, El-Boghdadly K, Burckett-St Laurent D. Artificial intelligence for image interpretation in ultrasound-guided regional anaesthesia. Anaesthesia. 2021;76(5):602-607. doi:https://doi.org/10.1111/anae.15212

Bowness J, Taylor A. Ultrasound-Guided Regional Anaesthesia: Visualising the Nerve and Needle. Adv Exp Med Biol. 2020 2020;1235:19-34. doi:doi:10.1007/978-3-030-37639-0_2

Bowness J, Varsou O, Turbitt L, Burkett-St Laurent D. Identifying anatomical structures on ultrasound: assistive artificial intelligence in ultrasound-guided regional anesthesia. Clinical Anatomy. 2021 2021;34(5):802-809. doi:doi:10.1002/ca.23742

Bowness JS, El-Boghdadly K, Woodworth G, Noble JA, Higham H, Burckett-St Laurent D. Exploring the utility of assistive artificial intelligence for ultrasound scanning in regional anesthesia. Regional Anesthesia & Pain Medicine. 2022:rapm-2021-103368. doi:10.1136/rapm-2021-103368

Champaneria R, Shah L, Wilson MJ, Daniels JP. Clinical effectiveness of transversus abdominis plane (TAP) blocks for pain relief after caesarean section: a meta-analysis. International Journal of Obstetric Anesthesia. 2016-12-1 2016;28:45-60. doi:doi:https://doi.org/10.1016/j.ijoa.2016.07.009

Cho C, Lee YH, Park J, Lee S. A self-spatial adaptive weighting based u-net for image segmentation. Electronics (Switzerland). 2021 2021;10(3):1-11. doi:doi:10.3390/electronics10030348

Connor CW. Artificial Intelligence and Machine Learning in Anesthesiology. Anesthesiology. 2019-12 2019;131(6):1346-1359. doi:doi:10.1097/aln.0000000000002694

Fayezizadeh M, Majumder A, Neupane R, Elliott HL, Novitsky YW. Efficacy of transversus abdominis plane block with liposomal bupivacaine during open abdominal wall reconstruction. The American Journal of Surgery. 2016-9-1 2016;212(3):399-405. doi:doi:https://doi.org/10.1016/j.amjsurg.2015.12.026

Feng-Ping A, Zhi-Wen L. Medical image segmentation algorithm based on feedback mechanism convolutional neural network. Biomedical Signal Processing and Control. 2019 2019;53doi:doi:10.1016/j.bspc.2019.101589

Fern, o T, Gammulle H, Denman S, Sridharan S, Fookes C. Deep Learning for Medical Anomaly Detection A Survey. ACM Computing Surveys. 2022 2022;54(7)doi:doi:10.1145/3464423

Goldsmith AJ, Liteplo A, Hayes BD, Duggan N, Huang C, Shokoohi H. Ultrasound-guided transgluteal sciatic nerve analgesia for refractory back pain in the ED: A case series. American Journal of Emergency Medicine. 2020 2020;38(9):1792-1795. doi:doi:10.1016/j.ajem.2020.06.001

Gungor I, Gunaydin B, Oktar SO, et al. A real-time anatomy ıdentification via tool based on artificial ıntelligence for ultrasound-guided peripheral nerve block procedures: an accuracy study. Journal of Anesthesia. 2021 2021;35(4):591-594. doi:doi:10.1007/s00540-021-02947-3

Hadjerci O, Hafiane A, Conte D, Makris P, Vieyres P, Delbos A. Computer-aided detection system for nerve identification using ultrasound images: A comparative study. Informatics in Medicine Unlocked. 2016 2016;3:29-43. doi:doi:10.1016/j.imu.2016.06.003

Hadjerci O, Hafiane A, Makris P, et al. Nerve Localization by Machine Learning Framework with New Feature Selection Algorithm. Springer International Publishing; 2015:246-256.

Hadjerci O, Hafiane A, Morette N, Novales C, Vieyres P, Delbos A. Assistive system based on nerve detection and needle navigation in ultrasound images for regional anesthesia. Expert Systems with Applications. 2016 2016;61:64-77. doi:doi:10.1016/j.eswa.2016.05.002

Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial Intelligence in Anesthesiology: Current Techniques, Clinical Applications, and Limitations. Anesthesiology. 2020-2 2020;132(2):379-394. doi:doi:10.1097/aln.0000000000002960

Henderson M, Dolan J. Challenges, solutions, and advances in ultrasound-guided regional anaesthesia. BJA Education. 2016;16(11):374-380. doi:10.1093/bjaed/mkw026

Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. Apr 8 2022;12(1):5979. doi:10.1038/s41598-022-09954-8

Sandhu NS. Ultrasound imaging in anesthesia: an overview of vascular access and peripheral nerve blocks. Seminars in Anesthesia, Perioperative Medicine and Pain. 2007-12-1 2007;26(4):197-209. doi:doi:https://doi.org/10.1053/j.sane.2007.10.003

Huang C, Zhou Y, Tan W, et al. Applying deep learning in recognizing the femoral nerve block region on ultrasound images. Annals of Translational Medicine. 2019 2019;7(18):30. doi:doi:

Jankovic Z, Ahmad N, Ravishankar N, Archer F. Transversus Abdominis Plane Block: How Safe is it? Anesthesia & Analgesia. 2008 2008;107(5)doi:doi:

Jeevakala S, Sreelakshmi C, Ram K, Rangasami R, Sivaprakasam M. Artificial intelligence in detection and segmentation of internal auditory canal and its nerves using deep learning techniques. International Journal of Computer Assisted Radiology and Surgery. 2020-11-1 2020;15(11):1859-1867. doi:doi:10.1007/s11548-020-02237-5

Jeng CL, Torrillo TM, Rosenblatt MA. Complications of peripheral nerve blocks. BJA: British Journal of Anaesthesia. 2010 2010;105:i97-i107. doi:doi:10.1093/bja/aeq273

Kavakli AS, Kavrut Öztürk N, Arslan Ü, Enginar F, Canim Ş, Uzunay E. The use of ultrasound guided combined peripheral nerve blocks in a high-risk patient: A case report. Agri. 2021 2021;33(1):39-41. doi:doi:10.5505/agri.2018.25902

Keller DS, Ermlich BO, Delaney CP. Demonstrating the Benefits of Transversus Abdominis Plane Blocks on Patient Outcomes in Laparoscopic Colorectal Surgery: Review of 200 Consecutive Cases. Journal of the American College of Surgeons. 2014-12-1 2014;219(6):1143-1148. doi:doi:https://doi.org/10.1016/j.jamcollsurg.2014.08.011

Koscielniak-Nielsen ZJ. Ultrasound-guided peripheral nerve blocks: What are the benefits? Acta Anaesthesiologica Scandinavica. 2008-7-1 2008;52(6):727-737. doi:doi:https://doi.org/10.1111/j.1399-6576.2008.01666.x

Laffey JG, Tobin É, Boylan JF, McShane AJ. Assessment of a simple artificial neural network for predicting residual neuromuscular block. British Journal of Anaesthesia. 2003 2003;90(1):48-52. doi:doi:10.1093/bja/aeg015

Lancaster P, Chadwick M. Liver trauma secondary to ultrasound-guided transversus abdominis plane block. British Journal of Anaesthesia. 2010-4-1 2010;104(4):509-510. doi:doi:https://doi.org/10.1093/bja/aeq046

Lei B, Huang S, Li H, et al. Self-co-attention neural network for anatomy segmentation in whole breast ultrasound. Medical Image Analysis. 2020 2020;64doi:doi:10.1016/j.media.2020.101753

Liu C, Liu F, Wang L, Ma L, Lu ZM. Segmentation of nerve on ultrasound images using deep adversarial network. International Journal of Innovative Computing, Information and Control. 2018 2018;14(1):53-64. doi:doi:10.24507/ijicic.14.01.53

Marhofer P, Greher M, Kapral S. Ultrasound guidance in regional anaesthesia†. British Journal of Anaesthesia. 2005;94(1):7-17. doi:10.1093/bja/aei002

Mariano ER, Marshall ZJ, Urman RD, Kaye AD. Ultrasound and its evolution in perioperative regional anesthesia and analgesia. Best Pract Res Clin Anaesthesiol. Mar 2014;28(1):29-39. doi:10.1016/j.bpa.2013.11.001

Marzola F, van Alfen N, Doorduin J, Meiburger KM. Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment. Computers in Biology and Medicine. 2021 2021;135doi:doi:10.1016/j.compbiomed.2021.104623

Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. npj Digital Medicine. 2020-9-24 2020;3(1):126. doi:doi:10.1038/s41746-020-00333-z

Mudumbai SC, Auyong DB, Memtsoudis SG, Mariano ER. A pragmatic approach to evaluating new techniques in regional anesthesia and acute pain medicine. Pain Management. 2018 2018;8(6):475-485. doi:doi:10.2217/pmt-2018-0017

Naidu RK, Richebe P. Probable Local Anesthetic Systemic Toxicity in a Postpartum Patient with Acute Fatty Liver of Pregnancy After a Transversus Abdominis Plane Block. A&A Practice. 2013 2013;1(5)doi:doi:

Niraj G, Kelkar A, Hart E, Kaushik V, Fleet D, Jameson J. Four quadrant transversus abdominis plane block and continuous transversus abdominis plane analgesia: a 3-year prospective audit in

124 patients. Journal of Clinical Anesthesia. 2015-11-1 2015;27(7):579-584. doi:doi:https://doi.org/10.1016/j.jclinane.2015.07.005

Pappada SM. Machine learning in medicine: It has arrived, let's embrace it. J Card Surg. 2021-11 2021;36(11):4121-4124. doi:doi:10.1111/jocs.15918

Peng SY, Peng SK. Predicting adverse outcomes of cardiac surgery with the application of artificial neural networks. Anaesthesia. 2008 2008;63(7):705-713. doi:doi:10.1111/j.1365-2044.2008.05478.x

Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. European radiology experimental. 2018 2018;2(1):35-35. doi:doi:10.1186/s41747-018-0061-6

Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Springer; 2015:234-241.

Salam AA, Aamir R, Khan RI, Ahmed A, Rehman A. Safety and effectiveness of ultrasound guided peripheral nerve blocks: Audit at tertiary care hospital. Journal of the Pakistan Medical Association. 2021 2021;71(6):1623-1626. doi:doi:10.47391/JPMA.447

Sammons G, Ritchey W. Use of Transversus Abdominis Plane (TAP) Blocks for Pain Management in Elderly Surgical Patients. AORN Journal. 2015-11-1 2015;102(5):493-497. doi:doi:https://doi.org/10.1016/j.aorn.2015.09.003

Santanen OAP, Svartling N, Haasio J, Paloheimo MPJ. Neural nets and prediction of the recovery rate from neuromuscular block. European Journal of Anaesthesiology. 2003 2003;20(2):87-92. doi:doi:10.1017/S0265021503000164

Saranteas T, Karakitsos D, Alevizou A, Poularas J, Kostopanagiotou G, Karabinis A. Limitations and Technical Considerations of Ultrasound-Guided Peripheral Nerve Blocks: Edema and Subcutaneous Air. Regional Anesthesia and Pain Medicine. 2008-7-1 2008;33(4):353-356. doi:doi:https://doi.org/10.1016/j.rapm.2007.12.013

Scherrer V, Compere V, Loisel C, Dureuil B. Cardiac Arrest from Local Anesthetic Toxicity After a Field Block and Transversus Abdominis Plane Block: A Consequence of Miscommunication Between the Anesthesiologist and Surgeon. A&A Practice. 2013 2013;1(5)doi:doi:

Shen X, Ma H, Liu R, Li H, He J, Wu X. Lesion segmentation in breast ultrasound images using the optimized marked watershed method. BioMedical Engineering Online. 2021 2021;20(1)doi:doi:10.1186/s12938-021-00891-7

Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. Mar 19 2019;19(1):64. doi:10.1186/s12874-019-0681-4

Sites BD, Chan VW, Neal JM, et al. The American Society of Regional Anesthesia and Pain Medicine and the European Society Of Regional Anaesthesia and Pain Therapy Joint Committee recommendations for education and training in ultrasound-guided regional anesthesia. Reg Anesth Pain Med. 2009-1 2009;34(1):40-6. doi:doi:10.1097/AAP.0b013e3181926779

Smistad E, Iversen DH, Leidig L, Lervik Bakeng JB, Johansen KF, Lindseth F. Automatic Segmentation and Probe Guidance for Real-Time Assistance of Ultrasound-Guided Femoral Nerve Blocks. Ultrasound in Medicine & Biology. 2017-1-1 2017;43(1):218-226. doi:doi:https://doi.org/10.1016/j.ultrasmedbio.2016.08.036

Smistad E, Johansen KF, Iversen DH, Reinertsen I. Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks. J Med Imaging (Bellingham). 2018-10 2018;5(4):044004. doi:doi:10.1117/1.JMI.5.4.044004

Smistad E, Østvik A, Pedersen A. High performance neural network inference, streaming, and visualization of medical images using FAST. IEEE Access. 2019 2019;7:136310-136321. doi:doi:10.1109/ACCESS.2019.2942441

Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for Reporting Machine Learning Analyses in Clinical Research. Circulation: Cardiovascular Quality and Outcomes. 2020;13(10):e006556. doi:doi:10.1161/CIRCOUTCOMES.120.006556

Tian D, Zhu B, Wang J, et al. Brachial Plexus Nerve Trunk Recognition From Ultrasound Images: A Comparative Study of Deep Learning Models. IEEE Access. 2022;10:82003-82014. doi:10.1109/ACCESS.2022.3196356

Tighe P, Laduzenski S, Edwards D, Ellis N, Boezaart AP, Aygtug H. Use of Machine Learning Theory to Predict the Need for Femoral Nerve Block Following ACL Repair. Pain Medicine. 2011 2011;12(10):1566-1575. doi:doi:10.1111/j.1526-4637.2011.01228.x

Tucker RV, Peterson WJ, Mink JT, et al. Defining an Ultrasound-guided Regional Anesthesia Curriculum for Emergency Medicine. AEM Education and Training. 2021 2021;5(3)doi:doi:10.1002/aet2.10557

Turbitt LR, Mariano ER, El-Boghdadly K. Future directions in regional anaesthesia: not just for the cognoscenti. Anaesthesia. Mar 2020;75(3):293-297. doi:10.1111/anae.14768

Wadhwa A, adai SK, Tongpresert S, Obal D, Gebhard RE. Ultrasound guidance for deep peripheral nerve blocks: a brief review. Anesthesiol Res Pract. 2011 2011;2011:262070. doi:doi:10.1155/2011/262070

Walker G. Transversus abdominis plane block: a note of caution! British Journal of Anaesthesia. 2010-2-1 2010;104(2):265. doi:doi:https://doi.org/10.1093/bja/aep387

Wassef M, Lee DY, Levine JL, et al. Feasibility and analgesic efficacy of the transversus abdominis plane block after single-port laparoscopy in patients having bariatric surgery. Journal of pain research. 2013 2013;6:837-841. doi:doi:10.2147/JPR.S50561

Weiss E, Jolly C, Dumoulin J-L, et al. Convulsions in 2 Patients After Bilateral Ultrasound-Guided Transversus Abdominis Plane Blocks for Cesarean Analgesia. Regional Anesthesia & Pain Medicine. 2014 2014;39(3):248. doi:doi:10.1097/AAP.0000000000000088

Weng Y, Zhou T, Li Y, Qiu X. NAS-Unet: Neural architecture search for medical image segmentation. IEEE Access. 2019 2019;7:44247-44257. doi:doi:10.1109/ACCESS.2019.2908991

Woodworth GE, Carney PA, Cohen JM, et al. Development and Validation of an Assessment of Regional Anesthesia Ultrasound Interpretation Skills. Reg Anesth Pain Med. 2015-7 2015;40(4):306-14. doi:doi:10.1097/aap.0000000000000236

Zhao H, Sun N, Zhao Y, Kong X, Taubman D. Improved U-Net Model for Nerve Segmentation. Springer International Publishing; 2017:496-504.

Zhou Z, Zhao G, Kijowski R, Liu F. Deep convolutional neural network for segmentation of knee joint anatomy. Magnetic Resonance in Medicine. 2018 2018;80(6):2759-2770. doi:doi:10.1002/mrm.27229